Proceedings of the 12th International Conference on Music Perception and Cognition and the 8th Triennial Conference of the European Society for the Cognitive Sciences of Music, July 23-28, 2012, Thessaloniki, Greece
Cambouropoulos E., Tsougras C., Mavromatis P., Pastiadis K. (Editors)

Voice Multiplicity Influences the Perception of Musical Emotions

Yuri Broze[*1] & Brandon Paul[#2]

[*]*School of Music, Ohio State University, USA*
[#]*Department of Speech and Hearing Science, Ohio State University, USA*
[1]broze.3@osu.edu, [2]paul.674@osu.edu

ABSTRACT

A polyphonic musical texture can be described in terms of its *voice multiplicity*—the number of simultaneous musical voices present. We conjectured that listeners might make use of voice multiplicity information when inferring the expression of musical emotions. In particular, we hypothesized that ratings of musical loneliness would be highest for monophonic music, and decrease as more voices are added to the texture. Moreover, voice multiplicity should only influence emotion perception to the extent that it can be accurately perceived. In an experimental study, listeners were asked to rate brief (5s) musical excerpts for expression of happiness, sadness, loneliness, and pride. We controlled for style, motivic content, timbre, and loudness by excerpting harpsichord recordings of fugue expositions from Bach's Well-Tempered Clavier. Higher loneliness and sadness ratings were associated with fewer musical voices; loneliness showed a stronger effect than sadness. The effect of voice multiplicity was consistent with the pattern predicted by limitations in stream segregation. Unexpectedly, listeners were much more likely to make strong emotion ratings for monophonic textures than for any other multiplicity level, and multiplicity effects seemed to be greater for loneliness and pride ratings than for sadness and happiness ratings. Preliminary results from a second study using an expanded between-groups design are consistent with the idea that positively-valenced emotions are more easily perceived when more musical voices are present, whereas negatively-valenced emotions are perceived more strongly when fewer voices are present.

I. INTRODUCTION

Musical textures can be classified in terms of the number of concurrent musical voices present at once—the music's *voice multiplicity*. Monophonic music consists of a single musical voice, while polyphonic music characteristically employs multiple simultaneous melodies. While many musical features which have been reliably associated with particular perceived musical emotions (for a recent review, see Gabrielsson and Lindström 2010), voice multiplicity appears not to have been systematically studied.

The attribution of animacy, personality, and emotional qualities to perceptual objects in the visual domain has been well-studied (Scholl and Tremoulet 2000). Heider and Simmel (1944) provided evidence of our propensity to describe nonsentient images—in this case simple geometric figures—as possessing human motivations, conflicts, and emotional states. In their classic study, subjects were shown an animated film in which circles and triangles moved about in ways suggestive of social interaction. Asked to respond to the scene, participants spontaneously described these shapes as animate beings with intentional and affective states.

Participants in studies such as this typically attribute emotional states to individual animated shapes, of which there might be several in view at once. This suggests the possibility that perceived musical emotions might in some cases be similarly oriented toward individual 'auditory objects'. If this is were true, then one might expect that emotions which imply particular social situations would be better expressed by music with an appropriate voice multiplicity.

We attempted to experimentally test the conjecture that the voice multiplicity of a musical passage could influence the perception of some musical emotions. In our first experiment, we asked participants to rate several brief musical excerpts for perceived loneliness, predicting that music with fewer musical voices would be perceived as more lonely.

We also hoped to distinguish effects caused by perception of voice multiplicity *per se* from those caused by possible differences in other musical features. For example, one would expect that as more voices are added to a musical texture, it would tend to show increases in dynamic level, textural density, or rhythmic complexity, among other features. A compelling demonstration of voice multiplicity effects on musical emotion perception would include a means to disentangle the effect of perceived voice multiplicity from that of other musical features.

In order to focus on voice multiplicity effects, we made use of previous research describing musical effects of certain perceptual limitations. Compared with visual objects, auditory objects are somewhat more difficult to accurately denumerate —that is, they are more difficult to distinguish and count. In the case of polyphonic music of homogenous timbre, it seems as though listeners are able to easily identify one-voice and two-voice textures, but easily confuse three-voice with four-voice textures (Huron 1989). This suggests that voice multiplicity effects on musical emotion perception might be similarly subject to certain perceptual constraints. Thus, we hoped to experimentally identify a pattern in perceptual loneliness which reflects these limitations.

To anticipate our results, we found that listeners do indeed associate loneliness with polyphonic textures of fewer musical voices. Moreover, these ratings followed a pattern consistent with the interpretation that it was voice multiplicity *per se* which gave rise to the observed effect. Moreover, we identified several other phenomena, prompting a follow-up study to be described.

II. EXPERIMENT ONE

A. Aims

In our first study, we wished to test the primary hypothesis that loneliness ratings would decrease as voice multiplicity increased. Moreover, we expected this effect to be caused by a perceptual awareness of the number of simultaneous musical voices in the musical texture.

Listeners can detect multiple concurrent musical voices in polyphonic textures, but stream segregation is subject to perceptual limitations. In particular, three-voice and four-voice textures are easily confused (Huron 1989). If the

number of concurrent musical voices of a polyphonic texture —the voice multiplicity—influences emotion perception, one might expect this effect to be limited by one's perceptual capacity to count them. Specifically, loneliness ratings should not differ substantially between three-voice and four-voice conditions. Hence, we also hypothesized that loneliness ratings would decline more between two- and three-voice conditions than between three- and four-voice conditions.

B. Stimuli

We wished to choose stimuli which would carefully control as many musical features as possible. 36 five-second fugal excerpts from J. S. Bach's *Well-Tempered Clavier* were collected, representing conditions of 1–4 concurrent voices. Fugues such as these typically begin with a single musical part presenting the melody which serves as the fugue's subject. Additional voices are subsequently added to the texture, each time repeating the same basic melodic content. By excerpting from fugue expositions, we generated sets of stimuli matched for both melodic content, tempo, and other composition-specific features. Because our principal interest was in perceptual loneliness (a negatively-valenced emotion), seven minor mode fugues (28 stimuli) were chosen for analysis. Additionally, two major-mode fugues (8 stimuli) were included to provide contrast.

All stimuli were taken from harpsichord recordings, as this instrument has a relatively uniform timbre across its range, as well as a restricted dynamic range. These instrumental characteristics would make it relatively difficult for harpsichordists to emphasize one musical part over the others using such cues as dynamics or articulation. By maximizing the independence of the musical voices, we hoped to reduce any confounding effects introduced by the performer, and facilitate the accurate perception of voice multiplicity.

C. Methods

25 Ohio State University School of Music students rated fully randomized excerpts for perceived sadness, happiness, loneliness, and pride for course credit. These additional emotions were chosen to reduce the risk of a demand characteristic while providing the opportunity for exploratory analysis of voice multiplicity influences on the perception of emotions other than loneliness. Happiness and sadness are prototypical positive and negative emotions. Pride was chosen as well, because, like loneliness, it carries specific connotations of social approval or disapproval, but is positively-valenced.

After 5 practice trials, participants rated all 144 stimulus/emotion combinations using a computer interface, followed by 72 randomly-selected stimulus/emotion pairs in a second block, for a total of 216 ratings. There did not appear to be a substantial effect of block, so ratings from both blocks were used in data analysis.

A slider interface was used to collect self-report data. The slider was labeled on the left and right side as "not at all [emotion]" (corresponding to a rating of 1.00) to "very [emotion]" (corresponding to 7.00). After each rating, the slider was reset to the center position. Although this measurement device does not typically provide normally-distributed data, it has the advantage of allowing subjects to indicate qualitative differences in their emotional perceptions, by deliberately moving the slider to an extreme position (1.00 or 7.00) or by leaving it untouched in the center (4.00).

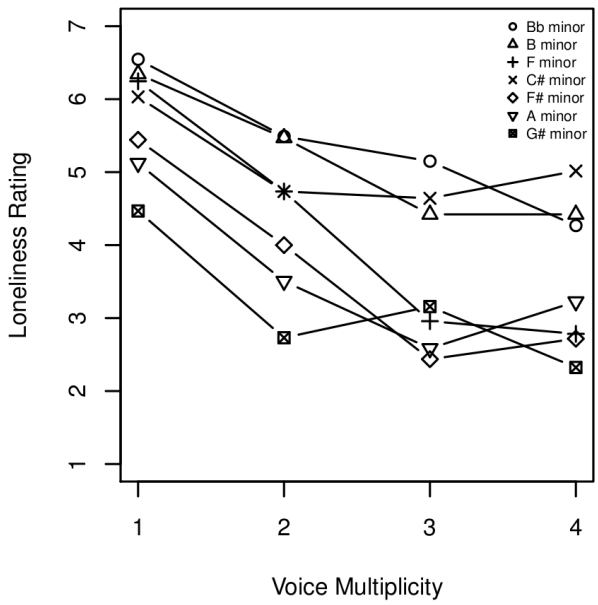

Figure 1. Median perceptual loneliness ratings for matched stimuli from each of seven fugues from J.S. Bach's *Well-Tempered Clavier* (N=25). There is a statistically significant decrease in perceived loneliness between one-voice and four-voice conditions (Fisher's sign test, $B_{193} = 162$, $p < .001$).

After participants made ratings, post-experimental interviews were conducted to explore rating strategies and check for demand characteristics.

D. Results

Our first hypothesis was that loneliness ratings would decrease as more voices were added to the texture. Median loneliness ratings for each minor fugue in the study are shown in Figure 1, plotted against voice multiplicity condition. Each stimulus set showed a net decrease in loneliness ratings as voice multiplicity increased. The results are consistent with the hypothesis that loneliness ratings will be higher for music with few voices and lower for music with many voices. As an inferential test, we conducted a one-sided Fisher sign test comparing the difference in rating per person per fugue between voice multiplicities of one and four, and found the difference to be statistically significant ($B_{193} = 162$, $p < .001$)

We were interested in exploring the relationship of polyphonic voice multiplicity with ratings for the positively-valenced social emotion pride, as well as with happiness and sadness ratings. Boxplots of the ratings are given in Figure 2. The positively-valenced emotions exhibited increasing ratings as voice multiplicity increased from 1 to 3, but did not appear to notably differ between 3-voice and 4-voice textures. Ratings for negatively-valenced emotions showed the opposite effect.

It also appeared that the effect of voice multiplicity on ratings for the social emotions (loneliness and pride) was stronger than for the nonsocial emotions (sadness and happiness). In a *post hoc* statistical test, we found that the difference in emotion ratings between one- and four-voice textures is greater for loneliness than for sadness ($z = -3.66$, $p < .001$).

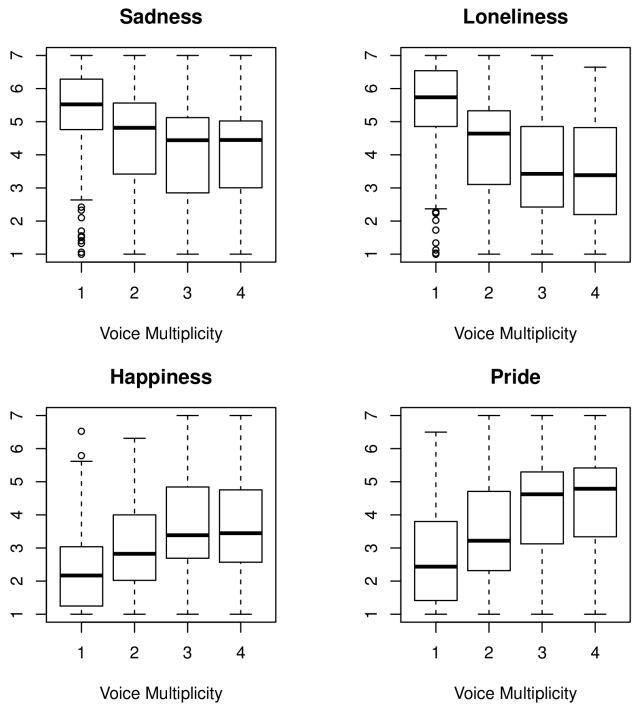

Figure 2. Ratings for the perception of two social and two nonsocial musical emotions for 5s excerpts from seven fugues (N = 25). Matched sets of stimuli were drawn from J.S. Bach's *Well-Tempered Clavier*. Positively-valenced emotions are rated higher for excerpts with more voices present; negatively-valenced emotions are rated lower. The pattern of ratings follows that predicted by limitations is voice denumerability (Huron 1989).

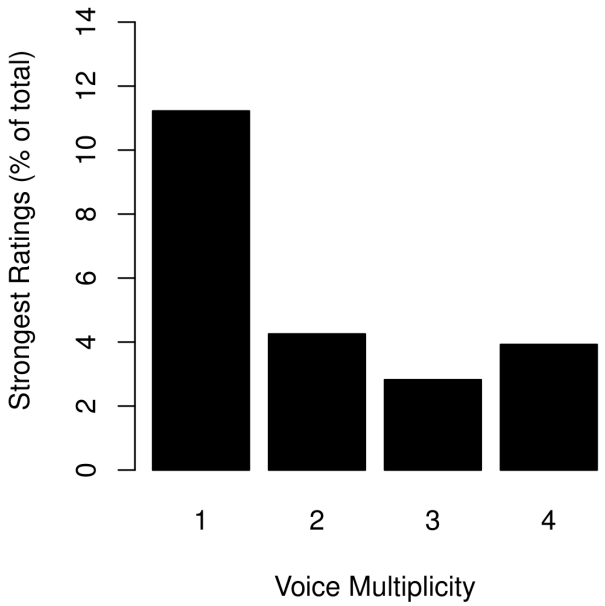

Figure 3. Strongest emotion ratings (1.0 or 7.0 on a continuous scale) as a percentage of total ratings made. Participants are more likely to make strong emotion ratings for monophonic music than polyphonic music (logistic $z = 10.59$, $p < .001$). Emotions rated were happiness, sadness, pride, and loneliness.

A *post hoc* exploratory review of the slider-based response data revealed that participants were much more likely to make ratings at the extreme values of the slider (1.00 or 7.00) for monophonic (1-voice) excerpts than for stimuli of any other voice multiplicity level (Figure 3).

E. Discussion

The results of our first experiment were consistent with the hypothesis that voice multiplicity influences perceived loneliness perception. Moreover, this effect appeared to reflect perceptual limitations on voice denumerability—voice multiplicity effects arose predominantly between one-, two-, and three-voice conditions.

We interpret *post hoc* analyses as suggesting that negatively-valenced emotions are more readily perceived in music with fewer voices, and positively-valenced emotions are more readily perceived in music with many voices. Additionally, it seemed that perception of social emotions (loneliness and pride) exhibited greater voice multiplicity effects than did perception of basic emotions (sadness and happiness). This is consistent with the idea that information about the number of voices present in a texture is particularly important in judging social emotion content.

Finally, *post hoc* analysis revealed that extreme emotion ratings were significantly more likely to occur when only a single voice is present in the music. If extreme slider ratings are interpreted as indicating high certainty of an emotion rating, then one could propose at least two possible accounts for the the observed effect. It could be that for the emotions tested, listeners interpreted monophonic music as depicting a single person, and took this to indicate a particular social setting. This information might have been used while making emotion judgments. Alternatively, it could be the case that perceived social situations played no role in the listener's ratings. In this case, one might suspect that melodic content plays an important role in the perception of some musical emotions, and that monophonic music simply presents this melodic information more clearly. In this second interpretation, multi-voice music is more confusing because it contains multiple simultaneously expressive melodies.

There are some potential confounds which should be acknowledged, including demand characteristics. In post-experimental interviews, textural density was very rarely mentioned as a strategy for making ratings, and there was no evidence that listeners were aware of the purpose of the experiment. However, because each listener was asked to make ratings for all four emotions, it is possible that the striking symmetry observed in the rating patterns was simply an artifact of the experimental design. That is, the ratings for different emotions were not entirely independent in the present design.

To address this issue, we attempted to replicate our results using a between-subjects design. This replicate study would allow us to substantiate some of our exploratory results.

III. EXPERIMENT TWO

A. Aims

In our second study, we aimed to replicate the results of our first study using a between-groups design. In addition to retesting the loneliness-related hypotheses of our first experiment, we made the following additional predictions: (1) Negatively valenced emotions would receive higher ratings when fewer voices were present. (2) Positively valenced emotions would receive higher ratings when more voices were present. (3) Social emotion ratings would be more strongly affected by voice multiplicity conditions than would nonsocial emotions. (4) Monophonic music would be more likely to receive an extreme rating (1.00 or 7.00 on a continuous scale) than music of any other voice multiplicity.

B. Stimuli and Methods

Stimuli used were identical to those employed in the first experiment. Because the major-mode and minor-mode excerpts from the first study did not appear to exhibit different response patterns, we decided to use all 36 stimuli (7 minor fugues and 2 major fugues, each with four voice multiplicity conditions).

Sophomore music students from the Ohio State University participated for course credit. Participants were assigned to either a positive-emotion or a negative-emotion group. Using the same slider interface as in the first study, the positive-emotion group rated the 'nonsocial' emotions of happiness, contentment, and excitement and the 'social' emotions of pride, love, and compassion; the negative-emotion group rated the 'nonsocial' emotions of sadness, fear, and disgust and the 'social' emotions of loneliness, envy, and shame. Each stimulus/emotion combination was presented in fully randomized order.

C. Preliminary Results

We are still in the process of collecting data. However, we are able to report some preliminary findings concerning the effect of voice multiplicity on positive and negative emotion perception. Boxplots of the emotion ratings obtained from the positive and negative emotion groups (n = 16 each) are presented in Figure 4. Negative emotions show a decrease and positive emotions show an increase in ratings between 1-voice and 4-voice conditions (Fisher's sign test $p < .001$). Additionally, we replicated the finding that extreme emotion ratings are significantly more likely when only a single voice is present in the musical texture (logistic $z = 7.535$, $p < .001$).

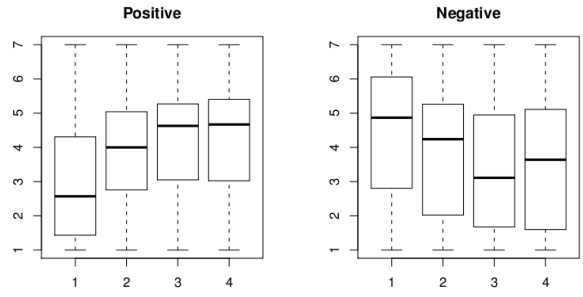

Figure 4. Preliminary results for Experiment Two. Median emotion ratings for the positive and negative emotion groups are given (n=16 each). Positive emotion ratings increase as multiplicity increases; negative emotion ratings decrease.

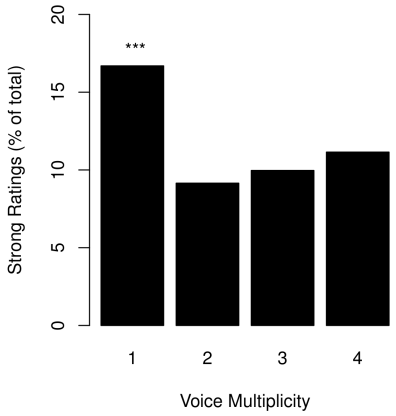

Figure 5. Preliminary results for Experiment Two. Monophonic excerpts are more likely to receive an extreme rating for perceived emotion than for any other voice multiplicity level (logistic $z = 7.535$, $p < .001$).

IV. DISCUSSION

The present study presents evidence consistent with the hypothesis that the voice multiplicity of music influences the perception of several musical emotions, including happiness, sadness, loneliness, and pride. It appears that the effect might generalize to positively and negatively-valenced emotions, with fewer musical voices associated with the perception of negative musical emotions. There is some evidence that suggests emotion perception might be qualitatively different between three different categories of music: Monophonic music, biphonic music, and polyphonic music of three or more voices. This is consistent with previously-described limitations on polyphonic voice denumerability (Huron 1989).

However, there are three caveats which bear mentioning. First, a musical interpretation of our results might point out that three theoretical constructions of western music could be associated with voice multiplicity conditions of one, two, or three voices: notes, intervals, and triads. It could be the case that listeners are responding differently to stimuli at these voice multiplicity levels due to the specific musical information which becomes available. This would suggest that it is not voice multiplicity *per se* which influences musical emotion perception, but instead the existence or nonexistence of notes, intervals, or chords. However, this interpretation cannot account for why loneliness ratings would exhibit greater sensitivity to voice multiplicity than sadness.

A second potential problem with this approach would be that we failed to directly account for several possible confounds inherent in the musical stimuli. It could very well be the case that listeners are responding to musical features other than voice multiplicity which have their own affective associations, such as melodic motion and rhythm. Indeed, associations have been found between perceived happiness and sadness with ascending and descending melodies respectively (Gerardi & Gerken 1995), and "firm" rhythms have been connected to perceived sadness while "flowing" rhythms have been associated with joyfulness (Wedin, 1972). Certainly these musical features vary within our sampled

musical stimuli. Work to better characterize our stimuli and formulate a model incorporating covariates is currently underway.

Finally, the generalizability of our results has not yet been established. One would wish to reproduce these results using other musical instruments and other genres of music. Moreover, because the effect of voice multiplicity appears to apply to positively and negatively valenced emotions quite generally, it would be desirable to fully account for the specific interactions voice multiplicity might have with musical mode in influencing musical emotion perception.

ACKNOWLEDGMENTS

The authors would like to thank Kate Guarna for her invaluable assistance collecting data, and CSML lab members for the helpful feedback.

REFERENCES

Gabrielsson, A. and Lindström, E. (2010). The Role of Structure in the Musical Expression of Emotions. *Handbook of Music and Emotion: Theory, Research, Applications.* Oxford: Oxford University Press.

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

Heider, F. and Simmel, M. (1944). An experimental study of apparent behavior. *American Journal of Psychology*, 57:243–259.

Huron, D. B. (1989). Voice denumerability in polyphonic music of homogeneous timbres. *Music Perception,* 6(4):361–382.

Scholl, B. J. and Tremoulet, P. D. (2000). Perceptual causality and animacy. *Trends in Cognitive Sciences*, 4(8).

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