THE EFFECT OF PITCH EXPOSURE ON SADNESS JUDGMENTS: AN ASSOCIATION BETWEEN SADNESS AND LOWER THAN NORMAL PITCH

David Huron, Gary Yim*, Parag Chordia†*

*School of Music, Ohio State University
†Department of Music, Georgia Institute of Technology

ABSTRACT

Participants heard melodies in an experiment consisting of an exposure phase and a test phase. During the exposure phase, pairs of participants heard melodies that used artificial scales: for one participant, certain scale tones were raised; for the other, those same scale tones were lowered. Then, during the test phase, each pair of participants judged the sadness of identical melodies without these alterations. Participants who had heard the exposure melodies with the raised scale tones heard the test melodies as lower, and judged the test melodies to be sadder. The results suggest that the minor mode achieves its sad qualia by contrast with a more familiar mode (major) in which some scale tones are lower than normal. The results also suggest why the minor scale may have no sad connotations for listeners not enculturated to the major scale.

1. INTRODUCTION

One of the various prosodic features associated with sad speech relates to the pitch of the voice. Fairbanks and Pronovost (1939), for example, found an association between sadness and low overall speaking pitch. In the case of music, Huron (2008) showed that instrumental themes in the minor mode are, on average, slightly lower in pitch than comparable major-mode themes, even after accounting for the lowered scale degrees in the minor mode.

However, the notion that sadness is linked to low pitch height is inconsistent with some commonplace experiences. For example, if low pitch were heard as "sad" then virtually all adult male voices would be expected to sound sad compared with female voices — which is not a widely reported experience. One possible explanation is that affective content is assessed only after the pitch is somehow normalized. For example, listeners may somehow infer the natural tessitura or compass of the voice, and then perceive the pitch as high or low compared with the inferred range. This raises the possibility that pitch-related affect is linked to relative rather than absolute pitch height: that is, one might conjecture that it is the perception of "relatively low" that leads to the assessment of sad affect. It is this hypothesis that we propose to test in the current study.

In brief, the experiment consists of exposing listeners to unfamiliar scales and melodies, and then subsequently asking them to judge the sadness of test melodies with some pitches altered. To anticipate our results, we will see that different affective responses can be induced when two listeners hear the identical test melody – depending on whether some of the constituent pitches are perceived as either lower or higher relative to pitches for that listener's immediately prior exposure melodies.

2. HYPOTHESIS

The motivating hypothesis for this study may be stated formally as follows: There is an association between pitch height and affect, where lower than normal pitches are judged as sadder.

3. METHOD

In brief, listeners judged the sadness of unfamiliar melodies played in unfamiliar tuning systems in an experiment consisting of two phases. Participants were paired, and a between-subjects design was used to test the experimental hypothesis.

3.1 Participants

Sixteen participants were recruited from the Ohio State University School of Music subject pool. Participants were undergraduate music majors, 7 males and 9 females. They were informed that the task would involve judging musical sadness in non-Western melodies.

3.2 Experimental Design

Many plausible influences might be at work in listener sadness judgments: loudness, tempo, interval size, familiarity, learned associations, resemblances to the minor scale, dissonant intervallic relationships, melodic contour, rhythmic patterns, etc. The best way to control for all of these factors is to use an identical stimulus, and manipulate only the immediately preceding listening experience. Thus, an important methodological feature of this work is the pairing of participants — a "between-subjects" design. The experiment was divided into two phases: an exposure phase (which was different for the paired participants), and a test phase (which was the same for the paired participants). Our goal was to have the two participants judge the sadness of identical melodies in the test phase differently, depending on what was heard during the exposure phase.

Participants were tested individually in an Industrial Acoustics Corporation sound attenuation room. They each heard 30 melodies in the exposure phase and another 30 melodies in the test phase. The melodies were short, with an average duration of about 30 seconds. Following each melody, a Likert scale appeared on a computer display, and sadness judgments were recorded.

3.3 Stimuli

The stimuli were constructed with the aim of subverting Western listening habits. They consisted of monophonic melodies based on Germanic folksongs transposed to exotic-sounding artificial scales and making use of an unusual timbre. Details of the stimuli follow below: first, the means of scale generation; second, the source of the melodies; third, the nature of the mapping of the scales to the melodies; and finally, a description of the timbre employed.

Scale Generation

For each pair of participants a single test scale (for the test phase) was generated. From this test scale, two modified exposure scales (for the exposure phase) were derived. Three criteria shaped the generation of each scale: they should be (1) exotic, (2) categorically unambiguous, and (3) average pitch equivalent.

Exotic. Since our participants were all enculturated to Western music, any resemblance of our stimuli to the major/minor scale system might be expected to confound the results. Accordingly, we aimed to generate scales that would subvert Western listening habits by avoiding familiar intervals, avoiding equal temperament, and avoiding octave equivalence. We did this simply by randomizing the distance (80 to 350 cents) between successive scale steps for each test scale. In order to avoid an excessively small or large pitch compass, a further condition was that the complete scale must lie between 800 and 1400 cents (i.e. between 8 and 14 semitones). All of the random test scales were asymmetric, with no fixed interval sizes. In randomly selecting scale tones for modification, the lowest scale tone was excluded from consideration for reasons explained below. Eight test scales were generated in the experiment: one 9-tone scale, one 8-tone scale, two 7-tone scales, two 6-tone scales, and two 5-tone scales.

Categorically unambiguous. For each test scale created, two exposure-phase scales were derived. A minimum of one tone was modified, with a maximum of one-third of the scale tones being modified. For the "high scale," the selected tones from the test scale were raised in pitch. For the "low scale" those same scale tones were lowered in pitch. The size of the pitch modification (in cents) was the same for both the "high" and "low" exposure-phase scales.

Our experimental design hinges on listeners hearing a modified pitch according to some pre-defined relationship. In order to ensure that modifications were perceptible, a minimum pitch modification of 50 cents (quarter tone) was established. This minimum distance is in the region of the perceptual limen for pitch interval judgments for Western listeners. Another crucial factor was ensuring that a given listener heard the deviant pitch as either "higher" or "lower" than normal. In order to minimize possible confusions as to which scale tone was modified, the modified pitch was always nearer to the pitch from which it was a nominal modification. Pitch proximity is known to have a powerful perceptual grouping effect, so it was reasonable to suppose that proximity to the nominal pitch would ensure the appropriate perception. The average pitch modification was 92 cents.

Average Pitch Equivalent. As manipulating the pitch will also change the average pitch height of the scale, a compensating transposition was performed. In practice, these were quite small (typically less than 30 cents), and were virtually undetectable to the experimenters.

Figure 1 summarizes the generation of the experimental scales.

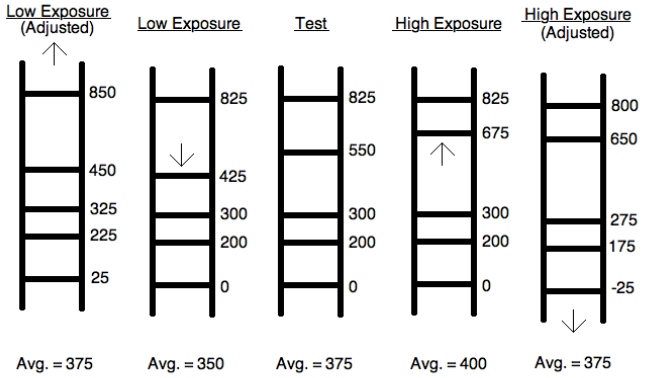

Figure 1: An illustration of the method of scale generation used in the experiment. Each ladder represents a different scale with rungs representing successive scale tones. Pitch distances are indicated in cents with zero representing some arbitrary pitch standard. A *test scale* is randomly generated with unequal step sizes (figure center). From this scale, two *exposure scales* are derived (next to center). In this example, a single scale tone has been either lowered or raised in order to create the "low" and "high" exposure scales. Since overall pitch height is known to influence judgments of affect, the modified exposure scales are then transposed slightly so that the average pitch height is identical for both scales.

Exposure Melodies

The melodies used in this study were based on unfamiliar Germanic folksongs drawn from the Essen Folksong Collection (Schaffrath 1995). Each folksong (numbering over six thousand) in this electronic collection is encoded in terms of diatonic scale degrees and chromatic alterations. Folksongs containing chromatic tones were excluded from the experiment. Each pair of participants heard a different set of melodies. The folksongs were mapped to the novel scales using a method described below.

Timbre

Throughout the experiment, our goal was to ensure that participants did not hear the melodies as out-of-tune variants of music in the Western major or minor scales. We assumed that unfamiliar timbres would contribute to the sense that the melodies originated in some exotic non-Western culture. Accordingly, synthesized timbres were modeled on the Kalimba (African thumb-piano). Despite the use of inharmonic timbres, the tones nevertheless evoked clear pitch perceptions.

Method of Melody-Scale Mapping

The folksongs melodies were mapped to the novel scale systems as follows. The number of unique pitches in a given folksong was

determined and a test scale with the same number of unique pitches was generated according to the criteria described above. The lowest pitch in the folksong was mapped to the lowest tone in the novel scale, the second lowest to the second lowest, and so on. This permitted a consistent one-to-one mapping between the pitches in the folksong and the novel scale.

For the purposes of this study, we designated the lowest pitch in all novel scales as the "tonic." The purpose of this was to facilitate the perception of a tonal center that might help listeners to hear the modified pitches as either lower or higher than the corresponding pitches heard in their respective exposure phases. In order to facilitate mapping of folksongs to these scales, we selected only those folksongs in which the tonic was the lowest pitch. Of the 6,255 folksongs in the database some 1,685 exhibited the tonic pitch as the lowest pitch in the melody: accordingly, we only selected our stimulus melodies from these.

In this way, we created "homologous" melodies in which the "tonic" pitches were the lowest pitches in both the source folksong melodies and in the mapped exotic-scale versions. This approach preserves whatever correlations exist between tonic pitches and rhythmic or phrase-related patterns in the original folksong. Thus, although the resultant melodies were clearly exotic-sounding, the melodic contours and rhythms of the original folksongs were preserved, such that the experimental stimuli sound like "real" melodies. Figure 2 below illustrates the principles involved in mapping folksong melodies to the novel scales.

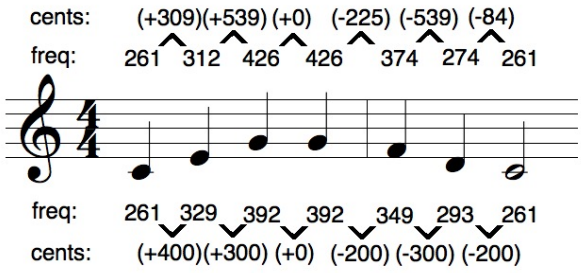

Figure 2: An illustration of the mapping between an original folksong excerpt and a novel scale. In this example, C4 is both the tonic and the lowest pitch in the original folksong. Values below the staff identify the fundamental frequencies (in Hz) for each tone and (in parentheses) the interval distances (in cents) according to equal temperament. Values above the staff identify the fundamental frequencies and interval distances for a novel melody created by mapping the randomly generated scale to this song.

4. RESULTS

The experiment was structured using a between-subjects design. In carrying out a between-subjects analysis, we begin by noting that some participants may tend to rate all melodies as generally sadder than other participants, and some participants may tend to respond using a narrower range of judgments. In order to match between-subjects variation, all of the raw judgments were normalized within subjects. For each subject we used that subject's mean and standard deviation for the exposure phase sadness ratings to normalize the test phase sadness ratings.

The differences in the normalized z-scores between the paired participants provides a way to characterize possible differences in emotional experiences when hearing the test stimuli. Figure 3 below summarizes the results for all eight pairs of participants.

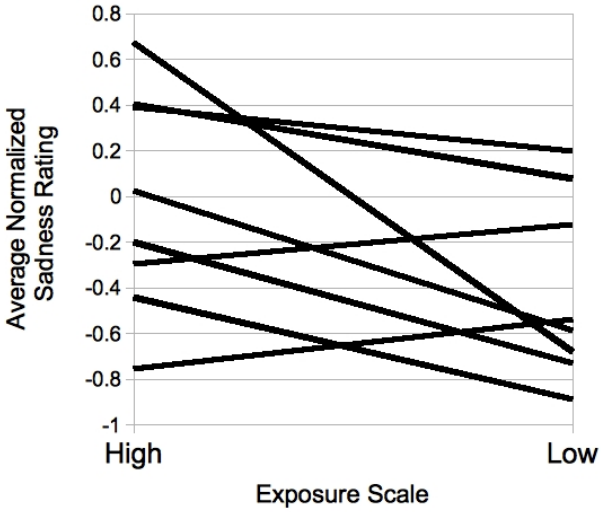

Figure 3: Summary of sadness judgments for eight pairs of participants. Plotted values represent the average normalized sadness ratings for the 15 test melodies for each participant. Values on the left of the diagram represent ratings for participants exposed to the "high" exposure scales (who heard the melodies as containing "lowered" pitches). Values on the right represent ratings for participants exposed to the "low" exposure scales (who heard the melodies as containing "raised" pitches). Lines join the values for paired participants hearing the identical test melodies. The majority of lines show a left-to-right descent: identical melodies were judged sadder on average when they were lower than normal (i.e., subsequent to high exposure), consistent with the experimental hypothesis.

A matched-pairs t-test was done with the 15 test melody judgments carried out by each of the 16 participants. That is, sadness judgments were matched for each identical melody heard by two participants — one of whom heard the melody as lower than than the exposure scale, while the other heard the melody as higher than the exposure scale. The mean difference in the z-scores across all pairs of participants was 0.38, in the direction consistent with the hypothesis. This difference is statistically significant ($t=3.315$, $df=119$, $p=0.0012$).

5. CONCLUSION

In this study, we tested the hypothesis that lower than normal pitch is associated with the perception of greater sadness. We aimed to demonstrate this by manipulating the sadness ratings of identical test melodies, given by listeners exposed for 15 minutes to different exposure melodies. One participant heard exposure

melodies using a “high scale” while the other heard the same melodies using a “low scale.” The results were that the identical test melodies were judged as sadder by the participant exposed to high scale melodies. Thus, if the tones heard during the exposure phase established a pitch “norm” for listeners, test melodies were heard as sadder if they were lower than normal. This is consistent with the experimental hypothesis.

There are some important caveats. First, it is important to emphasize that all of the participants in our experiment were undergraduate music students, principally from the midwest region of the United States, enculturated to the popular and art-music traditions as they exist in the early 21st century. The responses of the participants may not generalize to other listeners.

In addition, it is appropriate to note that in rendering their judgments of sadness, we made no effort to resolve the thorny aesthetic issue of whether nominally sad music either evokes or represents sadness. That is, we cannot distinguish whether listeners “felt” the melodies to evoke a sad experience, or whether they recognized the melodies as apt representations or expressions of sadness (see, e.g. Kivy 1980).

Finally, it should be noted that this experiment relied on a priming paradigm in which stimulus judgments were shown to depend upon immediately preceding stimuli. In discussing our results, we have suggested that this pitch difference might also account for the experience of sadness associated with the minor mode for Western-enculturated listeners. However, For Western-enculturated listeners, it does not seem necessary for listeners to hear an immediately preceding major-mode passage in order to experience a subsequent minor-mode passage as sounding sad. That is, our experience of the sadness of the minor mode may depend on a lifetime of exposure rather than any short-term priming effect. The current experiment has demonstrated an effect only at a short time interval.

6. DISCUSSION

Valentine (1913/1914) suggested that the purported sadness of the minor mode might arise as a conditioned response. That is, those sounds or sound patterns heard in sad contexts (such as memorial services) or accompanied by sad lyrics, would, after sufficient exposure, tend to evoke sad connotations on their own.

The results of this study are consistent with an alternative or complementary theory. The results suggest that the perception of sad affect arises from pitches that are lowered with respect to some learned pitch norm. In the case of Western music, if the major scale is regarded as a "default" or "norm" for Western-enculturated listeners, then the lowered pitches that are characteristic of the various minor scales might be reasonably construed as lower than normal. The notion that the perceived sadness in the minor mode may be caused by the modified pitches is consistent with the “lower than normal” principle implied by this experiment.

At the same time, the results do not suggest that there is anything inherently sad about the minor scales: the scales used in our experiment were created pseudo-randomly. Our experiment suggests that lowering arbitrary pitches has the hypothesized effect of increasing sadness judgments, implying that the minor scale accrues its sad connotation merely by contrast with a major-scale norm. The lower-than-normal principle may also account for the observation made by ethnomusicologists that the minor scale is not heard as sad in many cultures, especially those in the Middle East and in the Balkan region, where the Western major mode is not the preeminent or commonplace scale. If listeners are not enculturated to the major scale as "normative," there would be no reason to hear a minor scale as containing lower than normal pitches.

In light of this experiment, we may offer a tentative prediction. For any given musical culture, lowering one or more pitches from a commonplace or normative scale-like pitch collection might be expected to evoke sad connotations. This conjecture may provide a worthwhile topic for future research efforts.

7. REFERENCES

Fairbanks, Grant, and Wilbert Pronovost. 1939. An experimental study of the pitch characteristics of the voice during the expression of emotion. *Speech Monographs*, Vol. 6, pp. 87-104.

Heinlein, Christian Paul. 1928. The affective characteristics of the major and minor modes in music. *Journal of Comparative Psychology*, Vol. 8, pp. 101-142.

Hevner, Kate. 1935. The affective character of the major and minor modes in music. *American Journal of Psychology*, Vol. 47, pp. 103-118.

Huron, David. 2006. *Sweet Anticipation: Music and the Psychology of Expectation*. Cambridge, Massachusetts: MIT Press.

Huron, David 2008. A comparison of average pitch height and interval size in major- and minor-key themes: Evidence consistent with affect-related pitch prosody. *Empirical Musicology Review,* Vol. 3, No. 2, pp. 59-63.

Schaffrath, Helmuth. 1995. *The Essen Folksong Collection*. D. Huron (Ed.). Stanford, California: Center for Computer Assisted Research in the Humanities.

Valentine, Charles Wilfred. 1913/1914. The aesthetic appreciation of musical intervals among school children and adults. *British Journal of Psychiatry*, Vol. 6, pp.190-216.