Psychon Bull Rev (2015) 22:163–169
DOI 10.3758/s13423-014-0653-y

BRIEF REPORT

Melody recognition revisited: influence of melodic Gestalt on the encoding of relational pitch information

Yune-Sang Lee · Petr Janata · Carlton Frost · Zachary Martinez · Richard Granger

Published online: 28 May 2014
© Psychonomic Society, Inc. 2014

Abstract Melody recognition entails the encoding of pitch intervals between successive notes. While it has been shown that a whole melodic sequence is better encoded than the sum of its constituent intervals, the underlying reasons have remained opaque. Here, we compared listeners' accuracy in encoding the relative pitch distance between two notes (for example, C, E) of an interval to listeners accuracy under the following three modifications: (1) doubling the duration of each note (C – E –), (2) repetition of each note (C, C, E, E), and (3) adding a preceding note (G, C, E). Repeating (2) or adding an extra note (3) improved encoding of relative pitch distance when the melodic sequences were transposed to other keys, but lengthening the duration (1) did not improve encoding relative to the standard two-note interval sequences. Crucially, encoding accuracy was higher with the four-note sequences than with long two-note sequences despite the fact that sensory (pitch) information was held constant. We interpret the results to show that re-forming the Gestalts of two-note intervals into two-note "melodies" results in more accurate encoding of relational pitch information due to a richer structural context in which to embed the interval.

Y.-S. Lee (✉) · C. Frost · Z. Martinez · R. Granger
Department of Psychological and Brain Sciences, Dartmouth College, Hanover, NH 03755, USA
e-mail: yunesang.lee@gmail.com

P. Janata
Center for Mind and Brain, University of California, Davis, CA 95618, USA

Present Address:
Y.-S. Lee
Department of Neurology, University of Pennsylvania, 3 West Gates. 3400 Spruce st., Philadelphia, PA 19104, USA

Present Address:
C. Frost
Wisconsin Institutes for Medical Research II, 1111 Highland Ave., Madison, WI 53705, USA

Keywords Music · Melody · Gestalt · Interval · Pitch · Recognition

Introduction

Humans can easily recognize, reproduce, and remember melodies—sequences of musical notes. Except for extreme populations (e.g., listeners with absolute pitch or amusia), normal listeners recognize melodies based largely upon the relative sizes of the intervals between successive pitches, an ability still robustly preserved even when the entire frequency range of the music is shifted up or down (i.e., the key is changed). Relative pitch processing may be, if not innate, acquired early in development. For example, Plantinga and Trainor (2005) established that 5.5- to 6.5-month-old infants, after listening to a melody for 7 days, preferred to listen to a novel melody rather than the original melody, regardless of the key in which the original was played at test. Their subsequent experiment ruled out a possibility that the infants remembered absolute pitch information, suggesting that, like adults, their long-term representation of the melody was based on the sequence of relative-pitch intervals between tones. Electrophysiological studies have indicated that relative pitch interval processing occurs in an automatic fashion, even among non-musicians (Trainor et al., 2002). For example, Trainor et al. (2002) showed that an occasional deviation of ending note position (i.e., outside of the key) elicited a mismatch negativity (MMN)—a brain response commonly regarded as an indicator of automatic change detection in the absence of attention towards a stimulus.

Intriguingly, relative pitch information is better encoded when the length of the melodic sequence is increased to some extent, suggesting that greater melody note-count confers perceptual benefits in encoding relational pitch information. For example, Edworthy (1985) showed that recognizing a note that deviated from a key became easier as the number of notes in the melody increased. An influential theoretical model, the Krumhansl-Schmuckler key-finding algorithm, suggests that greater melody-note-count provides “tonal context,” which helps listeners establish the sense of key (Krumhansl, 1990). The key is an important framework that defines the relations between pitches that follow a particular musical rule (e.g., diatonic major scale) when they are transposed. Thus, tonality—the sense of key—is one of the primary factors that may contribute to efficient encoding of melodies.

There is evidence that adding one extra note to interval sequences is enough to improve relative pitch recognition (Cuddy & Cohen, 1976). Cuddy and Cohen initially tested the hypothesis that recognition of three-note melodies could be modeled as a linear combination of the recognition of individual two-note intervals constituting the sequence. Their findings, however, did not support this idea; instead, recognition of three-note intervals was substantially better than would be predicted by recognition of the constituent intervals, leading them to conclude that “there is a more effective encoding process to handle 3 note as opposed to 2 note (interval) test sequences… However, the question of whether or not the process involves the abstraction and synthesis of interval information is difficult to answer” (p. 264).

Although this finding can be explained well by Krumhansl and Shmuckler’s model, the observation also leaves open the possibility that other mechanisms, not yet fully explored, may come into play. Importantly, Cuddy and Cohen suggested a potential contribution of global structural context in encoding local interval processing of a melody. In their study, six different patterns of three-note melodies were used: Straight ascending, straight descending, and four versions in which there was a mixture of upward and downward intervals (Fig. 1a). Moreover, they further randomized the types of three-note sequences such that one of the three notes’ positions was either raised or lowered. This left a total of 36 different configurations of melodic sequences, each of which indeed elicited different degrees of recognizability. We combined data from musically untrained and trained participants in that study in order to better illustrate that recognition accuracies greatly varied across the 36 different patterns of three-note melodies (see Fig. 1a). For example, the accuracy was at its lowest (i.e., 0.45) when the middle note of the pattern 4 sequence was downward. By contrast, the accuracy was highest (i.e., 0.87) when the middle note of the pattern 2 sequence was upward. This hidden aspect of Cuddy and Cohen’s data indicates that the global configuration of melodic sequences, also known as ‘melodic Gestalt,’ may affect encoding of the pitch relation among the constituent notes. While the term gestalt is used primarily in the visual domain to reference holistic processing (a global percept that cannot be expressed solely as the sum of local parts), it also appears in the music/auditory literature. For example, Kubovy and Van Valkenburg (2001) discussed extensively how Gestalt processes may be at play in auditory object perception. In

A

pattern	1_{st} up	1_{st} dn	2_{nd} up	2_{nd} dn	3_{rd} up	3_{rd} dn	mean
1	0.68	0.80	0.64	0.80	0.78	0.69	0.73
2	0.66	0.67	0.60	0.87	0.78	0.63	0.70
3	0.61	0.74	0.60	0.69	0.64	0.67	0.66
4	0.60	0.65	0.45	0.82	0.75	0.55	0.64
5	0.50	0.73	0.64	0.63	0.65	0.73	0.65
6	0.61	0.67	0.58	0.74	0.51	0.66	0.63
mean	0.61	0.71	0.58	0.76	0.68	0.65	0.67

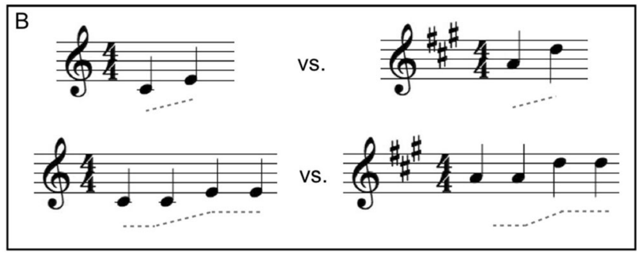

Fig. 1 **a** Recognition accuracy for various types of three-note sequences in Cuddy and Cohen (1976). The music notation depicts six different melodic patterns of three-note sequences. The bottom table summarizes the recognition accuracy when each of the six melodic patterns was manipulated, such that one of the constituent notes was either raised or lowered when the pattern was transposed to a different key. For example, the accuracy was 0.8 when pattern 1 was compared to its trasnsposition to other key with the first note raised by one semitone. **b** Schematic illustrations of the structural context effect in melody perception. *Top panel* A two-note interval comparison (e.g., major 3rd in C major vs perfect 4th in A major); *bottom panel* the same comparison with four notes. The *dotted-lines* illustrate the visual analogy of contour comparison. For the bottom pair, the *horizontal lines* on either side of the *sloping line* may provide a structural context to help distinguish the slope of the ascending contour

the music domain, Koelsch and Siebel (2005) used the term ‘melodic Gestalt formation’ when describing melody processing stages of their neurocognitive model of music perception, albeit without elaboration of mechanisms by which melodic gestalts might arise. Others have explored the idea of musical “Gestalts” more directly and formally (Neuhaus & Knösche, 2006; Schindler et al., 2012).

The present behavioral experiment sought to find direct evidence for Gestalt formation in the perception of relational pitch information associated with melodic intervals. Specifically, we compared the ability of listeners to encode four types of simple note sequences (Fig. 2a). In the first, baseline condition, the two pitches are played in a two-note sequence (such as C, E), each for 500 ms. In the second condition, the two notes are lengthened to 1,000 ms apiece, so that the same two-note sequence (C, E) is now 2 s, or twice its former duration. The third condition is identical duration to the second condition in duration (2 s), but each 500-ms note is repeated, i.e., two notes are substituted for each

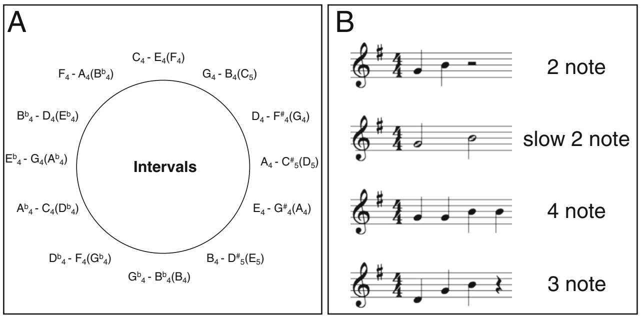

Fig. 2 **a** The entire set of intervals (depicted in circle of 5th format). Major 3rd or perfect 4th intervals in the reference keys of C and G (*bold font*) are compared to either of the same intervals transposed randomly to one of 12 keys. **b** Setup of note sequences (depicted in G key). Baseline intervals (two-note) are modified such that each note is played for a longer duration (long two-note) or is repeated (four-note). In another condition (three-note), a preceding note is added to the reference interval

of the longer ones (C, C, E, E). Finally, in the fourth condition, another note is prepended to the two-note sequences of condition 1 (G, C, E).

While the comparison between the two-note and three-note conditions (conditions 1 and 4, respectively) serves to corroborate the previous finding (Cuddy & Cohen, 1976), the primary theoretical significance of our study lies in the comparison of long two-note versus four-note (conditions 2 and 3). In particular, we expected to find an improvement in the encoding of the melodic patterns in the four-note condition over long two-note condition due to possible perceptual benefits associated with Gestalt (melodic patterns) change. For example, the added repeating notes might provide structural context to aid in disambiguating the subtle differences between intervals (Fig. 1b). Moreover, tonality could be established more strongly by reinforcing the tonal center, i.e., repeating the same two notes in the four-note condition. (Note that this is distinct from the key-finding algorithm in that no new tonal information is available.) By contrast, we expected to find no difference between the two-note and long two-note conditions because the perceptual quality is virtually identical (a single note followed by a second single note).

Materials and methods

Subjects

Thirteen right-handed participants (8 female; age range = 19–33 years, average 23.2 years) were recruited in the Dartmouth community. All participants reported that they previously received formal musical training on various types of musical instruments (musical training period = 1.5–8 years, average 4.6 years), although none of them were involved in any professional or semi-professional musical activities. Their hearing was normal as determined by responses to a questionnaire regarding hearing difficulties. None reported having perfect pitch. Consent forms were obtained from all listeners as approved by the Committee for the Protection of Human Subjects at Dartmouth College.

Stimuli

Figure 2b presents 24 intervallic sequences, arranged according to the circle of 5ths, which serve as the entire set of stimuli for the baseline two-note condition. These intervals also provide the basis for constructing the stimuli in each of three other conditions. The slow two- note condition was created by extending the duration (500 ms) of each note (i.e., sinusoidal frequency) to twice its length. The three-note condition was created by prepending a new note that was five seminotes lower (an interval of a perfect 4th) than the initial note of the target interval (i.e., the fifth scale degree in the octave below the initial note). Lastly, the four-note condition was created by repeating each constituent note of the interval. Figure 2a shows examples of each condition. Throughout the sequences, 20-ms linear ramps were applied to the onset and offset of each note to avoid acoustic transients.

Procedure

All listeners participated in each of the four experimental conditions except one listener who did not participate in the three-note condition. All four conditions were administered roughly 1–2 weeks apart from one another, and the order of the four conditions was fully randomized across subjects.

Trials were constructed as follows: a reference sequence was played, followed by a 2-s interstimulus interval (ISI), and then a target sequence followed by a 3.5-s response period. Listeners were instructed to press the mouse button corresponding to “same” if the relative intervals of the reference and target sequences were discerned to be identical, or the “different” button if they were perceived to be different. A terminal cue was presented to indicate the end of a trial and allow listeners to prepare for the next trial.

In all conditions, an ascending major 3rd or perfect 4th interval in one of two keys (C or G) were presented as reference sequences; then one of those two intervals in 1 of 12 keys were presented as target sequences, resulting in a total of 96 different comparisons, i.e., 2 (reference interval as major 3rd or perfect 4th) x 2 (reference interval in the key of C or G) x 2 (target interval as major 3rd or perfect 4th) x 12 (target interval in the key of C_4 to B_4).

In each experiment, the full set of 96 trials was presented twice with a 5–10 min break between sessions. A predetermined pseudo-random order of all trials was counterbalanced across listeners. No feedback was provided.

Results

Given the proportional data, we transformed the raw percent accuracy scores with the empirical logit prior to submitting them to standard linear statistics (Jaeger, 2008). We first conducted repeated-measures ANOVA, wherein the accuracies in the four conditions were entered as fixed effects. This analysis revealed a significant difference in performance across the conditions [$F(3,33)=3.16$, $P<0.05$)]. Then, we performed a series of planned contrasts using paired *t*-tests. The average recognition performance in each of the four conditions appears in Fig. 3. The first contrast revealed that performance was better with three-note than with two-note sequences [$t(11) = 2.32$, $P < 0.05$], replicating the previous finding by Cuddy and Cohen (1976). On the other hand, the second contrast revealed no difference between two-note and long two-note conditions [$t(12)= 1.54$, P=n.s.], indicating that increasing the sensory information did not help interval recognition. Lastly, and most importantly, the third contrast revealed that four-note melodies yielded higher encoding accuracy than did long two-note intervallic sequences [$t(12) = 2.27$, $P < 0.05$], indicating that changing the pitch sequence pattern from interval to melody improved performance despite holding the amount of sensory information constant.

We next conducted identical statistical comparisons using d-prime scores (Fig. 3), which yielded results consistent with those described above: (1) repeated ANOVA revealed an overall difference among the four conditions [$F(3,33)=3.09$, $P<0.05$]; (2) three-note vs two-note condition [$t(11)=2.27$, $P<0.05$]; (3) long two-note vs two-note [$t(12)=1.42$, P=n.s.]; (4) four-note vs two-note condition [$t(12)=2.21$, $P<0.05$].

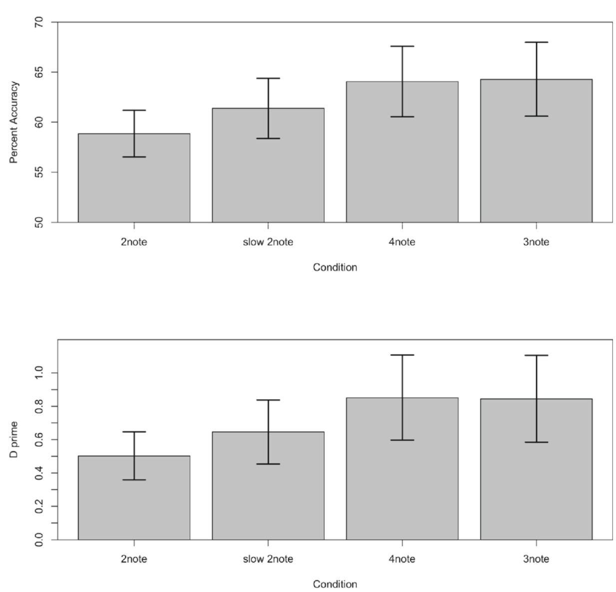

Fig. 3 *Top* Bar plot of mean accuracies across four different conditions; *bottom* bar plot of d prime across four different conditions. For both metrics, there were significant differences between two-note vs three-note, and long two-note vs four-note, whereas there was no difference between standard and long two-note intervals. *Error bars* Standard error of the mean

Additionally, we sought to examine the effects of key distance (i.e., six different key distances based upon circle of 5ths), interval type (either major 3rd or perfect 4th), reference key (i.e., C vs G major key), and trial type (i.e., whether the target sequence interval matched that of the reference sequence) on recognition performance across the four conditions. To this end, we ran a logistic mixed effects model via *lme4* (version 1.0-5) in R (version 3.0.2) with the binomial distribution (Bates et al., 2013; Winter, 2013). As fixed effects, all categorical variables (e.g., experimental condition, interval type, reference key, and trial type) were dummy coded except for the key distance variable which was centered since the variable was treated as continuous (i.e., from distance 1 to 6). As random effects, we included random intercepts for subjects. The analysis revealed a main effect of trial type with higher accuracy for "same" than for "different" reference and target sequences [$\beta= 0.8$; SE=0.11; z statistics= 7.4; $P <0.05$] , as well as a main effect of reference key with higher accuracy for C major than for G major key [$\beta= 0.45$; SE=0.11; z statistics= 4.12; $P <0.05$]. Although there was no overall main effect of interval type [$\beta= 0.01$; SE=0.11; z statistics= 0.1; P=n.s.], there was a significant interaction between the 3 note condition and interval type [$\beta= 0.48$; SE=0.13; z statistics= 3.77; $P <0.05$]. Lastly, we did not find evidence of key distance effect in the present study [$\beta= 0.03$; SE=0.04; z statistics= 0.7; P=n.s.].

Discussion

The present study sought to determine whether the mechanisms for encoding of pitch interval information differ between sequences consisting of only two notes and those consisting of more than two notes, i.e. whether an "interval-based" mechanism and a "melody-based" mechanism, respectively, can be disambiguated. While our results are in line with previous reports (Cuddy & Cohen, 1976; Dowling, 1986; Edworthy, 1985), an important aspect of the present finding is that transforming the Gestalt of two-note "intervallic" sequences into that of four-note "melodic" sequences did, in fact, also yield better encoding of relational pitch information, despite holding the total amount of sensory pitch information constant across those conditions.

This result provides novel evidence that melody recognition benefits not only from tonal context (Dowling, 1986) and duration (Dowling et al., 2008), but also from changes in melodic structure. Below, we review some potential mechanistic processes underlying this apparent change in melodic "Gestalt" that may lead to the enhanced encoding.

Presence of structural context

While long two-note and four-note sequences were matched almost exactly in their pitch-based sensory information, the main difference between the two sequences is the complexity of the contour. The long two-note condition consists of a single ascending segment, whereas the same ascending contour (e.g., C-E) is bracketed by the preceding (e.g., C-C-) and succeeding (e.g., -E-E) 'horizontal' contours in the four-note condition (Fig. 1b). As we surmised, the relative pitch distance of the interval to be judged (C2-E3) is represented better within such a structural context, as manifested in improved detection of changes in that pitch distance.

This effect may be understood by analogy to processing of similar forms in the visual domain. A briefly flashed line segment is distinguished better from other lure lines differing in orientation or location when those segments form part of a unitary object than when the line is presented alone (Enns & Prinzmetal, 1984; Mcclelland, 1978; Williams & Weisstein, 1978) (Fig. 1b). Related observations in the visual domain include holistic face processing, whereby differences in local details (e.g., the distance between eyes) are better detected when the whole face is shown than when the image of the face is modified, such that the top and bottom halves of the face are misaligned (Richler et al., 2011).

This interpretation accords with evidence suggesting that global contour processing influences local pitch processing (Stewart et al., 2008), and that melodic context facilitates pitch constancy judgments (Warrier & Zatorre, 2002). It has been hypothesized that more detailed analysis of pitch relations among constituent notes may be *preceded* by melodic Gestalt formation (Koelsch & Siebel, 2005). In line with the evidence from both auditory and visual experiments, we find that fine-grained details of local relationships can be better analyzed when additional global context is provided. The additional context transforms the Gestalt of an interval into the Gestalt of a melodic fragment.

Tonality (presence of harmonic context)

In the present study, major third and perfect fourth intervals were transposed into 12 different keys. Because the difference between these two intervals is only a single seminote (the smallest intervallic step in Western tonal music), and because both of these intervals are perceived as consonant intervals in Western tonal music, they could be discerned as being similar rather than different when transposed. Such propensity was revealed by a response bias, wherein listeners tended to judge the pitch sequences in comparisons as being the same more often than as being different. This bias may have led to the trial type effect across the four conditions. However, the fact that d' was higher in both three-note and four-note conditions than in the baseline two-note condition suggested that the 'sameness' bias was reduced by changing the melodic Gestalt.

Alternatively, could such modifications be viewed as reinforcing tonality: that is, 'a sense of key'? According to Krumhansl's key finding algorithm (Krumhansl, 1990), either adding tonal notes or increasing duration can provide an important cue for establishing tonality. In particular, it is quite conceivable that the improved performance in the four-note condition is due to pitch-based entrainment (in what follows, we separately discuss the possibility of temporal entrainment). This conjecture stems from observations that sensory priming may account for relatedness judgments and speeded reaction times associated with notes or chords that follow a tonally related context (Leman, 2000; Bigand et al., 2003). Thus, the first pitch of the four-note melodic sequence may serve as a 'tonic primer' (i.e., the four-note melody is interpreted as beginning with the tonic, the most stable note in a key), which in turn may help generating expectations for the upcoming notes within the same key (see Collins et al., 2014, for a discussion of sensory vs cognitive priming in music perception). This notion is supported further by the observation that prepending another tonal note (e.g., 5th) to major 3rd (e.g., D-G-B) yielded better performance than to perfect 4th (e.g., D-G-C), since the former has stronger sense of tonal center than the latter.

Similarly, the tonal benefit in the four-note condition can be also accounted for by Deutsch's model (Deutsch, 1969). According to the model, onset patterns for notes forming each interval generate the units that underlie pitch-distance and contour abstraction within a whole melodic sequence. In the case of the two-note sequence, the listener is given information from a single pitch interval to draw on. However, when each note is presented twice, the listener has six times more interval information to draw on within the same key: i.e., in our current four-note setup, there are a total of six pair-wise combinations among the notes $C_1 - C_2$; $C_1 - E_1$; C_1–E_2; $C_2 - E_1$; $C_2 - E_2$; $E_1 - E_2$ (Note that the subscript '1' and '2 ' refer to the order of presentation of each pitch, not frequency range on the keyboard). Thus, these repeating notes may reinforce the tonal center, resulting in better encoding of pitch relations within the same key without introducing yet another note from the key (as in the case of the three-note condition).

Duration

Although we have discussed the possible effect of structural and tonal context on the encoding of pitch distance, we also note that the current findings do not strictly rule out other possible mechanisms by which the Gestalt is transformed from an interval to a melodic fragment. While not reaching significance, there was a trend of higher accuracy and d prime in the long two-note condition, compared to the two-note condition. This suggests that increased duration of sensory information may, to some extent, help in analyzing the pitch distance of transposed intervals. Consistent with the current observation, Dowling et al. (2008) reported a partial effect of duration (i.e., slow tempo), such that increasing duration (i.e., slowing down tempo) improved performance only on familiar tunes. Presumably, a short fragment of melodic sequence (e.g., C-C-G-G, the beginning of 'twinkle twinkle little star'), rather than a simple interval (e.g., C-G) tends to be stored in the long-term memory. Because our interval stimuli do not appear to qualify as 'familiar tunes,' given the non-uniqueness of two-note intervals, they are less likely to benefit from an increase in duration.

Familiarity

Related to the notion above, the particular four-note sequences employed in the current study (e.g., C-C-E-E and C-C-F-F) do commonly appear in music tunes, and it is therefore possible that they are stored in long-term memory. Thus, when a reference sequence of a familiar four-note melody was heard, listeners might have unconsciously invoked and retrieved the familiar tunes, which could be compared immediately to a target sequence of another familiar four-note tune. Once again, even if long two-note and four-note sequences were matched in the sensory information, the perceptual quality and the degree of familiarity may differ between the two sequences.

Attentional mechanism

Lastly, we consider attentional facilitation. According to this account, the first note summons attention to the onset of the following note sequence and may facilitate interval recognition afterwards. This interpretation, based on a theory of temporal attending (Jones, 1976; Jones et al., 2002) derives from evidence that an isochronous sequence of context notes increases the accuracy of judgments of whether a probe pitch is the same as the standard presented at the beginning of the sequence or if it is a seminote higher or lower. Although temporal parameters were not manipulated in the present study, we were able to test the hypothesis using Cuddy and Cohen's dataset. In that study, target melodies were formed by shifting either the first, second or third note upward or downward. We examined non-musicians' performance on trials wherein either the first or third note was altered. If the temporal entrainment account holds true, performance should improve when the third note is altered because the third note can benefit from an already established temporal context. In other words, the first note would simultaneously serve as an alerting note that would both reinforce the identical second note and allow a temporal context to be established. However, we failed to find evidence to support this hypothesis, as there was no significant difference between first and third note-altered melodies [$t(11)= 0.5$, n.s]. Although the possibility of attentional facilitation through temporal entrainment is in need of more rigorous testing to be fully discounted, it does not appear at the present time to suffice for explaining the improved recognition of three-note or four-note melodic fragments relative to a simple two-note interval.

Conclusion

We demonstrated experimentally that the relational pitch information is encoded more robustly in sequences consisting of more than two note events, even when only the same notes are used, i.e., when relevant sensory information is matched (as in the long two-note condition). We believe these results demonstrate that such a change of Gestalt from interval to melodic fragment yields better recognition of transposed pitch sequences due to more elaborate structural context. Intriguingly, following the experiment almost all listeners reported that the four-note condition was "more musical" than either of the two-note conditions. This suggests that changes in simple auditory Gestalts are associated with subjective phenomenological differences experienced by listeners. That is, a four-note sequence increases rhythmic and contour information over both standard and long two-note intervals. Although the addition of notes to a small initial number of notes may trivially result in a more musical-sounding sequence, our observations suggest that the sense of musicality may be a reflection on the functioning of a sequencing mechanism that links together successive intervals. The poor recognition of transposed intervals played in isolation indicates that their access to this representational mechanism is limited, even though it is sequences of such intervals that come to be represented by this mechanism. Thus, the integration of intervals into a single melody Gestalt appears to facilitate the accurate encoding of the pitches of the notes making up the melodies. We acknowledge that further research is needed to corroborate our notion at both the behavioral and the neural level. For example, a recent fMRI study found invariance in the pattern of neural activity in the auditory cortex in response to the same melodic Gestalt rendered in different keys or by different instruments (Schindler et al., 2012).

Acknowledgments The authors would like to thank Cory Kendrick, Kevin Miller, and Samuel Lloyd for their great help on data collection. We thank Bodo Winter for providing the helpful tutorial on the linear mixed effects modeling in R and his advice on the analysis for our study via personal communication with us. Yune-Sang Lee's special thanks go to Prof. Jay Hull for his enormous and unconditional help on other statistical analyses. Lastly, we are truly grateful to the reviewing editor—Dr. Bob McMurray—and two anonymous reviewers for their great comments and suggestions.

References

Bates, D., Maechler, M., Bolker. B., Walker S. (2013). lme4: Linear mixed-effects models using Eigen and S4. R package version 1.0-5. http://CRAN.R-project.org/package=lme4

Bigand, E., Poulin, B., Tillmann, B., Madurell, F., & D'Adamo, D. A. (2003). Sensory versus cognitive components in harmonic priming. *Journal of Experimental Psychology-Human Perception and Performance, 29*(1), 159–171. doi:10.1037/0096-1523.29.1.159

Collins, T., Tillmann, B., Barrett, F. S., Delbe, C., & Janata, P. (2014). A combined model of sensory and cognitive representations underlying tonal expectations in music: From audio signals to behavior. *Psychological Review, 121*(1), 33–65. doi:10.1037/a0034695

Cuddy, L. L., & Cohen, A. J. (1976). Recognition of transposed melodic sequences. *Quarterly Journal of Experimental Psychology, 28*(May), 255–270.

Deutsch, D. (1969). Music recognition. *Psychological Review, 76*(3), 300–307.

Dowling, W. J. (1986). Context effects on melody recognition - scale-step versus interval representations. *Music Perception, 3*(3), 281–296.

Dowling, W. J., Bartlett, J. C., Halpern, A. R., & Andrews, M. W. (2008). Melody recognition at fast and slow tempos: Effects of age, experience, and familiarity. *Perception & Psychophysics, 70*(3), 496–502. doi:10.3758/Pp.70.3.496

Edworthy, J. (1985). Interval and contour in melody processing. *Music Perception, 2*(3), 375–388.

Enns, J. T., & Prinzmetal, W. (1984). The role of redundancy in the object-line effect. *Perception & Psychophysics, 35*(1), 22–32.

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards Logit Mixed Models. *Journal of Memory and Language, 59*(4), 434–446. doi:10.1016/j.jml.2007.11.007

Jones, M. R. (1976). Time, our lost dimension - toward a new theory of perception, attention, and memory. *Psychological Review, 83*(5), 323–355.

Jones, M. R., Moynihan, H., MacKenzie, N., & Puente, J. (2002). Temporal aspects of stimulus-driven attending in dynamic arrays. *Psychological Science, 13*(4), 313–319.

Koelsch, S., & Siebel, W. A. (2005). Towards a neural basis of music perception. *Trends in Cognitive Sciences, 9*(12), 578–584. doi:10.1016/j.tics.2005.10.001

Krumhansl, C. L. (1990). *Cognitive foundations of musical pitch*. New York: Oxford University Press.

Kubovy, M., & Van Valkenburg, D. (2001). Auditory and visual objects. [Research Support, U.S. Gov't, P.H.S. Review]. *Cognition, 80*(1-2), 97–126.

Leman, M. (2000). An auditory model of the role of short-term memory in probe-tone ratings. *Music Perception, 17*(4), 481–509.

Mcclelland, J. L. (1978). Perception and masking of wholes and parts. *Journal of Experimental Psychology-Human Perception and Performance, 4*(2), 210–223.

Neuhaus, C., & Knösche, T. R. (2006). Processing of rhythmic and melodic gestalts—An ERP study. *Music Perception, 24*(2), 209–222. doi:10.1525/Mp.2006.24.2.209

Plantinga, J., & Trainor, L. J. (2005). Memory for melody: Infants use a relative pitch code. *Cognition, 98*(1), 1–11. doi:10.1016/J.Cognition.2004.09.008

Richler, J. J., Cheung, O. S., & Gauthier, I. (2011). Holistic processing predicts face recognition. [Research Support, N.I.H., Extramural Research Support, U.S. Gov't, Non-P.H.S.]. *Psychological Science, 22*(4), 464–471. doi:10.1177/0956797611401753

Schindler, A., Herdener, M., & Bartels, A. (2012). Coding of melodic Gestalt in human auditory cortex. *Cerebral Cortex*. doi:10.1093/cercor/bhs289

Stewart, L., Overath, T., Warren, J. D., Foxton, J. M., & Griffiths, T. D. (2008). fMRI evidence for a cortical hierarchy of pitch pattern processing. [Research Support, Non-U.S. Gov't]. *PloS One, 3*(1), e1470. doi:10.1371/journal.pone.0001470

Trainor, L. J., McDonald, K. L., & Alain, C. (2002). Automatic and controlled processing of melodic contour and interval information measured by electrical brain activity. *Journal of Cognitive Neuroscience, 14*(3), 430–442. doi:10.1162/089892902317361949

Warrier, C. M., & Zatorre, R. J. (2002). Influence of tonal context and timbral variation on perception of pitch. [Research Support, Non-U.S. Gov't]. *Perception & Psychophysics, 64*(2), 198–207.

Williams, A., & Weisstein, N. (1978). Line segments are perceived better in a coherent context than alone - object-line effect in visual-perception. *Memory & Cognition, 6*(2), 85–90.

Winter, B. (2013). Linear models and linear mixed effects models in R with linguistic applications. arXiv:1308.5499. [http://arxiv.org/pdf/1308.5499.pdf]