Recognition of transposed melodies: Effects of pitch distance and harmonic distance

Abigail L. Kleinsmith¹ · W. Trammell Neill¹

Published online: 27 November 2017
© Psychonomic Society, Inc. 2017

Abstract

People easily recognize a familiar melody in a previously unheard key, but they also retain some key-specific information. Does the recognition of a transposed melody depend on either pitch distance or harmonic distance from the initially learned instances? Previous research has shown a stronger effect of pitch closeness than of harmonic similarity, but did not directly test for an additional effect of the latter variable. In the present experiment, we familiarized participants with a simple eight-note melody in two different keys (C and D) and then tested their ability to discriminate the target melody from foils in other keys. The transpositions included were to the keys of C# (close in pitch height, but harmonically distant), G (more distant in pitch, but harmonically close), and F# (more distant in pitch and harmonically distant). Across participants, the transpositions to F# and G were either higher or lower than the initially trained melodies, so that their average pitch distances from C and D were equated. A signal detection theory analysis confirmed that discriminability (d') was better for targets and foils that were close in pitch distance to the studied exemplars. Harmonic similarity had no effect on discriminability, but it did affect response bias (c), in that harmonic similarity to the studied exemplars increased both hits and false alarms. Thus, both pitch distance and harmonic distance affect the recognition of transposed melodies, but with dissociable effects on discrimination and response bias.

Keywords Music cognition · Sound recognition · Signal detection theory

The recognition of familiar melodies presents a paradox: On the one hand, people easily recognize a familiar melody, even if it is transposed to a key in which they have not heard it before (e.g., Bartlett & Dowling, 1980). For melodies transposed to another key, the pitch of every note is altered, but the psychoacoustic changes between pitches (*intervals*) are preserved. Accordingly, theories of music recognition have proposed that melodies are represented in memory more abstractly than the immediately perceived notes or pitches—likely, as a sequence of intervals or an overall “contour” (Deutsch, 1969, 1972; Dowling, 1978; Dowling & Bartlett, 1981).

On the other hand, some research has suggested that people do retain pitch-specific information about melodies (e.g., Halpern, 1989; Levitin, 1994; Schellenberg & Trehub,

2003). As a particularly compelling example, Schellenberg and Trehub (2003) tested participants on their ability to distinguish instrumental themes from popular television shows from the same themes transposed one or two semitones above or below the original. Even for themes shifted one semitone apart, participants were above chance in distinguishing between the original themes and their transpositions.

Although some studies have provided evidence that pitch-specific information is retained in memory, it is less clear that such information affects the identification of melodies encountered in a new key. For example, we do not know whether the transpositions of TV themes in Schellenberg and Trehub (2003) were more difficult to identify than the original themes. It is possible that melody identification depends only on an abstract representation, whereas additional pitch-specific information only contributes to distinguishing a new instance of the melody from a previous one.

One study that has directly addressed this question was conducted by Schellenberg, Stalinski, and Marks (2014), with a follow-up study by Schellenberg and Habashi (2015). In Schellenberg et al. (2014), participants initially heard versions of unfamiliar folk melodies with median pitches of G₄ or C#₅ (six semitones apart). At test, participants were required to

✉ Abigail L. Kleinsmith
akleinsmith@albany.edu

✉ W. Trammell Neill
wneill@albany.edu

¹ Department of Psychology, University at Albany, State University of New York, Albany, NY 12222, USA

discriminate the target melodies from the new melodies, with half of the target melodies being transposed to the opposite key. Although the transposed melodies were still recognized well, their mean confidence ratings were lower than those for melodies in the original key.

Given that a key change is detrimental to melody recognition, the question arises of whether the decrement depends on the distance of the new key from the original one. The “distance” between keys can be conceptualized as either the change in overall pitch height or the harmonic relationship between keys. The latter is a music-theoretic construct based on the “cycle of fifths,” but it can be thought of as the number of notes shared in common by the two keys. Van Egmond, Povel, and Maris (1996) found that both kinds of distance affect the perceived similarity of transpositions of a melody, although distance in pitch height had a stronger effect.

A thesis by Kleinsmith (2015) revealed a much stronger effect of pitch distance than of harmonic similarity on the recognition of transposed melodies. Participants were familiarized with a simple eight-note melody in the keys of both C and D. They were then tested on their ability to discriminate the target melody from similar foils, either in the original keys or transposed to the key of C# or G (in the latter case, either higher or lower in pitch). It may be noted that C# is very close in pitch height to both C and D (one semitone from each), but it is harmonically distant from them. In contrast, G is farther away in pitch height (five semitones above D or below C), but it is harmonically close to C and D (neighboring both on the cycle of fifths). Target discrimination was much better in the harmonically unrelated, but closer in pitch, key of C# than in the harmonically related key of G. Additional experiments replicated this dominance of pitch distance over harmonic distance in affecting the recognition of transpositions.

Although pitch distance appears to have a stronger effect on the recognition of transposed melodies, the results of Kleinsmith (2015) did not rule out an additional effect of harmonic similarity. For example, in the experiment cited above, recognition in the harmonically related key of G might still be better (or worse) than in an unrelated key, if pitch distance could somehow be equated.

Bartlett and Dowling (1980) concluded that harmonic distance does in fact affect short-term recognition judgments. In their experiments, participants judged whether two successive melodies were the same or different. The participants rejected foils in distant keys more easily than foils in near keys. Curiously, “same” judgments to targets were not affected significantly by key distance. Although the authors did not compute a joint measure of discriminability such as d' (e.g., Macmillan & Creelman, 2005; Wickens, 2002), this pattern suggests that the discriminability of targets from foils was actually worse for near than for distant keys. Because Bartlett and Dowling’s task did not require a long-term representation of the target melody, it remains an open question

whether similar results would be obtained for identification based on long-term memory.

In the present experiment, we first familiarized participants with a simple eight-note target melody in two different keys (C and D). Participants were then given practice at discriminating the targets from similarly constructed foils. A final, test phase included the target melodies and foils transposed to the keys of C#, F#, and G. Although a melody in F# is close in pitch height to a melody in G, the key of F# (unlike G) is harmonically distant from the trained keys of C and D. Moreover, across participants, we varied whether the transpositions to F# and G were higher or lower than the trained keys of C and D. It may be noted that in the higher register, G is one semitone farther than F# from C and D, but in the lower register, G is one semitone closer than F#. Thus, *on the average*, the transpositions to F# and G were equated for pitch distance, while differing markedly in harmonic similarity to C and D. As such, this experiment provided a direct test of the effect of harmonic distance, as well as a replication of the effect of pitch distance (C# vs. F# or G).¹

Method

Participants

Participants were recruited from introductory psychology courses at the University at Albany, for which they received participation credit. A short demographic questionnaire was given prior to the experiment itself, to assess musical background. Of the 96 participants, 23 indicated having had some formal music training. Participants were tested individually in a quiet room, in a single session lasting approximately 30 min.

Stimuli and apparatus

Six eight-note monophonic melodies were created using Finale Notepad, free software designed for music writing and music notation (Fig. 1). The timbre was “Grand Piano,” and the duration of each note was 500 ms; hence, each melody had a duration of 4 s. All of the notes in a melody were consistent with the key defined by the first and last notes of the melody (e.g., the key of C in Fig. 1). In an attempt to make discrimination of the target from the foils adequately challenging, all six melodies had the same first three and last two notes, differing only in the middle three notes. The melodies were diatonic—that is, the notes were consistent with the key

¹ In principle, the experiment could be conducted with initial training in only one key. However, using two keys allowed us to roughly equate C# and F# on their average harmonic distances from C and D (see the Results section). More important, we expected this procedure to facilitate participants’ understanding that a melody could be the “same” despite a change in key (i.e., to wholly different notes).

Fig. 1 The six monophonic, eight-note melodies used as stimuli in all phases in this report, shown here in the key of C.

defined by the first and last notes of the melody. Transpositions of each of the melodies were created in the keys of C, C#, D, F#, and G. For half of the participants, the transpositions to F# were in the register higher than C and D (i.e., beginning on the pitch F#₄), and the transpositions to G were in the lower register (i.e., beginning on the pitch G₃). This pattern was reversed for the other half of the participants (i.e., the melodies began on the pitches F#₃ and G₄).²

The E-Prime E-studio software, version 1.2 (Psychology Software Tools, Pittsburgh, PA), running on a Dell Optiplex 745 computer, was used to control stimulus presentation and to collect response data. The stimuli were presented at a comfortable level on Panasonic RP-HT21 headphones. Participants responded “yes” or “no” by pressing the “c” or “,” key on the computer keyboard, respectively.

Procedure

The experiment was conducted in three phases. In the first phase, participants were told that they would hear an excerpt from a folk melody called “The Silver Stream” and that they would hear it in two different keys. For each participant, one of the six melodies in Fig. 1 was selected as the target (with the

rest reserved for recognition foils). The participant listened to the selected target melody ten times—five times in the key of C and five in D—with instructions to remember the melody. A 4-min distractor task, requiring the addition of two-digit numbers, preceded the next phase.

In the second phase, participants practiced discriminating the target melody in C and in D from foil melodies in the keys of C, C#, D, F#, and G. The target was presented ten times in C and ten in D, randomly intermixed with the five foil melodies, presented once in each of the five keys. Participants were instructed to respond “yes” or “no” on each trial, to the question “Is this a version of the song you studied?” It was emphasized that they should respond “yes” to the target melody irrespective of its starting note (i.e., its key). The question prompt remained visible on each trial until the participant had responded. Visual feedback, in the form of the word “correct” or “incorrect,” appeared for 1 s, followed by a 500-ms blank interval before presentation of the next melody. After completing the second phase, participants again engaged in a 4-min distractor task.

The final, test phase was similar to the second phase, except that the target melodies were now presented in the keys of C#, F#, and G, in addition to the studied keys of C and D. The target melody was presented five times in each of the five keys, randomly intermixed with the five foil melodies presented once in each of the five keys. As such, there were a total of 50 test trials, half targets and half foils.

Results

Figure 2 displays the proportions of correct responses to target melodies (hits) and foil melodies (correct rejections) for each key. In addition to these two measures, we computed the composite signal detection theory measures of discriminability (d') and response bias (c) for each condition, by participants. Each measure was subjected to a 2×5 mixed-model analysis of variance, with the between-subjects variable of group (F#₄/G₃ vs. F#₃/G₄) and the within-subjects variable of key (C, C#, D, F#, and G). Hit rates, d' , and c all differed significantly between keys: for hits, $F(4, 376) = 17.65$, $MSE = .026$; for d' , $F(4, 376) = 12.02$, $MSE = 0.959$; for c , $F(4, 376) = 10.33$, $MSE = .328$ (all $ps < .0005$). Correct rejections, however, did not differ significantly across keys, $F(4, 376) = 1.27$, $p = .281$, $MSE = .021$.

The variable of group did not significantly interact with key in any of the measures of interest (all $ps > .170$), indicating that the higher or lower pitch height of F# or G relative to the trained keys (C and D) had a negligible effect on recognizing transpositions. Accordingly, for the pairwise comparisons reported below, we collapsed over the two counterbalancing groups.

² The higher and lower transpositions to F# and G were thus split between participants, to allow comparisons to the other new key, C#, without confounding repetitions of a given key in either the specific pitch height or the pitch class (independent of height).

Fig. 2 Mean proportions of correct hits (dark gray bars) and correct rejections (light gray bars) as a function of key. (The means for F# and G are collapsed over the higher and lower registers.)

Pitch-distance effects (C# vs. F#)

The key of C# is the closest one possible in pitch distance to C and D (one semitone away from each), but it is harmonically distant from both, sharing only two notes in common (for musicians, five steps away on the cycle of fifths). The key of F# is maximally distant from C harmonically, sharing only one note (six steps away on the cycle of fifths), and it is considered distant from D as well, sharing only three notes (four steps away on the cycle of fifths). Although we do not know the exact psychophysical function relating harmonic distance to perceptual difference, the average harmonic distances of C# and F# from C and D are at least roughly equated. As such, the comparison of C# and F# is a relatively pure test of the pitch-distance effect. C# yielded more hits than F# (.88 vs. .75), $t(95) = 4.87$, $p < .0005$, but did not differ significantly in correct rejections (.79 vs. .82), $t(95) = -1.63$, $p = .106$. Accordingly, discriminability was higher for C# than for F# ($d' = 2.80$ vs. 2.37), $t(95) = 3.09$, $p = .003$, and the measure of response bias was lower ($c = -.187$ vs. $.169$), $t(95) = -4.07$, $p < .0005$. (All reported t tests are two-tailed.)

Harmonic-distance effects (F# vs. G)

The key of G is harmonically close to C and D, differing from each by only one constituent note (i.e., for musicians, one step from each on the cycle of fifths). The key of F# is harmonically the most distant key from C (six steps away) and is relatively distant from D (four steps away). By transposing to the register above or below the trained keys of C and D, the average pitch distance from the trained keys (five

semitones above D or below C) was equated. As such, the comparison of F# to G constitutes a relatively pure test of the harmonic-distance effect. The hit rate for G (.81) was significantly higher than that for F# (.75), $t(95) = 2.33$, $p = .022$, but the correct rejection rate (.79 vs. .82) approached significance in the opposite direction, $t(95) = -1.93$, $p = .056$. This trade-off between hits and correct rejections resulted in a significant difference in the measures of response bias ($c = -.062$ vs. $.169$), $t(95) = 2.83$, $p = .006$, whereas the difference in discriminability ($d' = 2.49$ vs. 2.37) was nullified, $t(95) = 0.62$, $p = .538$.

Discussion

In the present experiment, participants were first familiarized with a short melody in two different keys (C and D). They were then tested on their ability to discriminate transposed target melodies from foil melodies in the keys of C#, F#, and G, as well as in the original keys. A melody in C# is close in pitch height to C and D (one semitone from both) but is harmonically distant from them. In contrast, a melody in G is harmonically close to C and D but is more distant in pitch height (five semitones above D or below C). Finally, F# is harmonically distant from both C and D, and here was equated to G in average pitch distance (four semitones above D or six semitones below C). Replicating previous experiments (Kleinsmith, 2015), discrimination was strongly affected by pitch distance, regardless of harmonic relatedness. Recognition in the harmonically unrelated key of C# was

nearly as good as in the trained keys of C and D, and was superior to recognition in the keys of F# and G.

Although the hit rate was somewhat higher in G (harmonically close) than in F# (harmonically distant), this was offset by a decrease in correct rejections (i.e., more false alarms), resulting in a change in the signal detection measure of response bias, c , rather than in discriminability (d').³ In other words, harmonic relatedness results in both targets and foils being perceived as more similar to the studied melodies. Although the difference in false alarm rates between G and F# was small here, the shift in response criterion is consistent with the increase in false alarm rates with harmonic similarity found by Bartlett and Dowling (1980) when pairs of melodies were judged as either “same” or “different.”

That the hit rate in C# was substantially higher than the correct rejection rate, yielding an even greater difference in c than we observed for F#, might be construed as a similar effect of pitch distance on response bias. However, this result should be interpreted with caution, given that pitch distance clearly affected discriminability. Although the signal detection theory measure d' is usually assumed to be independent of response bias, measures of response bias (typically c or the alternative measure β or $\log \beta$) are generally *not* independent of discriminability. Without going into detail, measures of response bias reflect the distance of a response criterion from the midpoint between the presumed noise and signal-plus-noise distributions (see Macmillan & Creelman, 2005; Wickens, 2002). Any change in d' necessarily shifts that midpoint and so changes the placement of a fixed response criterion relative to that midpoint.

Our findings that pitch distance affects discriminability (and maybe response bias) and that harmonic distance affects response bias are both consistent with the general proposition that key- or pitch-specific information is retained in memory for melodies (e.g., Halpern, 1989; Levitin, 1994; Schellenberg & Trehub, 2003). If melodies were retained in a purely abstract form, such as a sequence of intervals, or a contour (Deutsch, 1969, 1972; Dowling, 1978; Dowling & Bartlett, 1981), neither effect would be obtained. It is not immediately obvious *why* harmonic similarity affects response bias. Because harmonically close keys share more notes in common, it is possible that participants are influenced by the repetition of familiar pitches, irrespective of their placement in the melody. Alternatively, it is possible that participants have some intuitive understanding not only of keys (or tonal centers), but also of the harmonic relations between them. This is plausible, particularly because changes in key within a piece

of music (i.e., modulations) are typically to a neighbor on the “cycle of fifths.”

With regard to the effect of pitch distance on discriminability, the present data do not distinguish between the distance from the studied exemplars themselves and the distance from a common representation abstracted from them. That is, in the present experiment, the new key of C# was not only close in pitch height to the studied exemplars in C and D, it was in fact the psychophysical *average* of the two pitch heights. Posner and Keele (1968, 1970) proposed that a category prototype is learned by averaging over variations on a pattern. Recognition of the pattern then depends on distance from that average (see also Rips, Shoben, & Smith, 1973; Rosch, 1975; Rosch, Mervis, Gray, Johnson, & Boyes-Braem, 1976). It is conceivable that when a melody is encountered in different keys, what is retained in memory is a prototype of the melody in an “average” key. This remains an issue for further investigation.

Finally, the question of generalizability can be raised with regard to *any* set of music materials selected for study. In the present experiment the melodic sequences were short and simple, consistent with the instructions to the participants that the target was excerpted from a folk melody, “The Silver Stream.”⁴ Much music—most classical, and some popular—is more complex, with wider variation in note durations and pitches, not to mention in dynamics and rhythmic structure. It might not be possible to implement the present experimental manipulations with more complex materials or with melodies that are already highly familiar, without maintaining rigorous methodological control. Imagine, for example, perturbing certain notes in “The Star Spangled Banner”: Discrimination would probably be so accurate as to preclude any meaningful test of distance effects. There would be other cues, such as the distinctive intervallic structure of certain phrases, that would guarantee discrimination of the target from otherwise similar distractors, regardless of the transposition. However, the present experiments clearly demonstrated that both pitch distance and harmonic distance have effects on the recognition of simple, recently learned melodic phrases. If melody recognition depends on the acquisition of an entirely abstract (pitch- and/or key-independent) mental representation, this would surely be easier for simple melodies than for more complex ones. Therefore, we confidently conclude that both pitch distance and harmonic distance do play roles in the learned recognition of melodies.

³ In signal detection theory analysis, a change in response criterion away from $c = 0$ (or $\beta = 1$) adds more errors than correct responses. Hence, it is expected that the decrease in hits for F#, relative to G, would be accompanied by a smaller increase in correct rejections.

⁴ Musicians who can sight-read can easily verify from Fig. 1 that the stimuli are plausibly excerpted from folk melodies. As anecdotal support, one of the research assistants, a classically trained musician, was unaware that the materials were specifically created for the study. He complained of an “earworm” (an involuntary, intrusive recurrence of the auditory memory), and he searched the Internet in vain for the complete rendering of “The Silver Stream.”

Author note These data were initially presented at the 15th Annual Auditory Perception, Cognition, and Action Meeting, in Boston, Massachusetts, November 2016. We are grateful to Jeff Bostwick for composition of the stimuli; to Nini Niniashvili, Diana Rumpf, Jessica Simon, and Destiny Valentine for assistance with data collection; and to George Seror, Enzo Belli, and Ron Friedman for helpful discussion of this work.

References

- Bartlett, J. C., & Dowling, W. J. (1980). Recognition of transposed melodies: A key–distance effect in developmental perspective. *Journal of Experimental Psychology: Human Perception and Performance*, 6, 501–515. doi:<https://doi.org/10.1037/0096-1523.6.3.501>
- Deutsch, D. (1969). Music recognition. *Psychological Review*, 76, 300–307. doi:<https://doi.org/10.1037/h0027237>
- Deutsch, D. (1972). Octave generalization and tune recognition. *Perception & Psychophysics*, 11, 411–412. doi:<https://doi.org/10.3758/bf03206280>
- Dowling, W. J. (1978). Scale and contour: Two components of a theory of memory for melodies. *Psychological Review*, 85, 341–354. doi:<https://doi.org/10.1037/0033-295x.85.4.341>
- Dowling, W. J., & Bartlett, J. C. (1981). The importance of interval information in long-term memory for melodies. *Psychomusicology*, 1, 30–49. doi:<https://doi.org/10.1037/h0094275>
- Halpern, A. R. (1989). Memory for the absolute pitch of familiar songs. *Memory & Cognition*, 17, 572–581. doi:<https://doi.org/10.3758/bf03197080>
- Kleinsmith, A. L. (2015). *Key generalization of recognition memory for melodies* (Unpublished master's thesis). University at Albany, State University of New York, Department of Psychology.
- Levitin, D. J. (1994). Absolute memory for musical pitch: Evidence from the production of learned melodies. *Perception & Psychophysics*, 56, 414–423. doi:<https://doi.org/10.3758/bf03206733>
- Macmillan, N. A., & Creelman, C. D. (2005). *Detection theory: A user's guide* (2nd ed.). Mahwah: Erlbaum.
- Posner, M. I., & Keele, S. W. (1968). On the genesis of abstract ideas. *Journal of Experimental Psychology*, 77, 353–363. doi:<https://doi.org/10.1037/h0025953>
- Posner, M. I., & Keele, S. W. (1970). Retention of abstract ideas. *Journal of Experimental Psychology*, 83, 304–308. doi:<https://doi.org/10.1037/h0028558>
- Rips, L. J., Shoben, E. J., & Smith, E. E. (1973). Semantic distance and the verification of semantic relations. *Journal of Verbal Learning and Verbal Behavior*, 14, 665–681. doi:[https://doi.org/10.1016/s0022-5371\(73\)80056-8](https://doi.org/10.1016/s0022-5371(73)80056-8)
- Rosch, E. (1975). Cognitive representations of semantic categories. *Journal of Experimental Psychology: General*, 104, 192–233. doi:<https://doi.org/10.1037/0096-3445.104.3.192>
- Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology*, 8, 382–439. doi:[https://doi.org/10.1016/0010-0285\(76\)90013-x](https://doi.org/10.1016/0010-0285(76)90013-x)
- Schellenberg, E. G., & Habashi, P. (2015). Remembering the melody and timbre, forgetting the key and tempo. *Memory & Cognition*, 43, 1021–1031. doi:<https://doi.org/10.3758/s13421-015-0519-1>
- Schellenberg, E. G., Stalinski, S. M., & Marks, B. M. (2014). Memory for surface features of unfamiliar melodies: Independent effects of changes in pitch and tempo. *Psychological Research*, 78, 84–95. doi:<https://doi.org/10.1007/s00426-013-0483-y>
- Schellenberg, E. G., & Trehub, S. E. (2003). Good pitch memory is widespread. *Psychological Science*, 14, 262–266. doi:<https://doi.org/10.1111/1467-9280.03432>
- Van Egmond, E., Povel, D. J., & Maris, E. (1996). The influence of height and key on the perceptual similarity of transposed melodies. *Perception & Psychophysics*, 58, 1252–1259.
- Wickens, T. D. (2002). *Elementary signal detection theory*. New York: Oxford University Press.