# Effects of Familiarity, Key Membership, and Interval Size on Perceiving Wrong Notes in Melodies

Rachna Raman,*1 Kevin Herndon,*2 W. Jay Dowling*3

*School of Behavioral & Brain Sciences, The University of Texas at Dallas, USA

1rachna.raman@utdallas.edu, 2kxh140630@utdallas.edu, 3jdowling@utdallas.edu

## ABSTRACT

Previous investigations showed that Western participants' perception of wrong notes in familiar Western melodies was influenced primarily by key membership (diatonic or nondiatonic) and to a lesser extent by interval size (1 or 2 semitones away from the original note). These results were supported by a cross-cultural study between South Indian classical (Carnātic) and Western musicians and nonmusicians with highly familiar Western melodies. However, Indian participants were slower and less accurate with Carnātic than with Western melodies, presumably due to the complexity of the Carnātic music system. In this study, we examined the effect of song familiarity in Westerners' perception of wrong notes. We chose 32 Western melodies previously rated as familiar, and 32 highly unfamiliar melodies, similar in style to the familiar melodies. Participants heard each melody twice, each time with one wrong note that was determined from one of eight possible types of wrong note based on key membership, interval size, and direction (up vs. down). Participants identified the wrong note by pressing a key. The results indicated an effect of music experience with unfamiliar melodies only, with musicians detecting wrong notes faster than nonmusicians. All groups were faster with familiar melodies than with unfamiliar melodies. Key membership influenced perception of wrong notes in both familiar and unfamiliar melodies: Participants were slower at recognizing wrong notes that were diatonic and faster when they were nondiatonic. Interval size influenced perception of wrong notes only in unfamiliar melodies: Participants were slower at recognizing wrong notes that were 2 semitones away than when they were 1 semitone away, which was the opposite of our previous studies with familiar melodies. We take these results as converging evidence that people remember the pitches in melodies in terms of steps of an overlearned modal scale, and not as successive intervals.

## INTRODUCTION

The melodic-rhythmic contour of a melody—the pattern of ups and downs of pitch along with the relative durations of the notes—is an important feature, both in terms of musical structure and of memory. But there is more to a melody than just contour; the pitch intervals among the notes are also important. (Consider the contrast between the tune "Twinkle, Twinkle" and Haydn's famously surprising theme.) Hence the question arises: are those intervals represented in memory as such, or as patterns of pitch classes (musical scales)?

Dowling (1978) proposed that melodies are represented in memory by two components: contour and scale. The contour represents the pattern of ups and downs, along with relative rhythmic values. Since most melodic motion is in diatonic steps, stepwise motion is "unmarked" in the representation, whereas leaps are "marked" and their size in scale steps specified. When a melody is produced from memory, the contour is "hung" on the appropriate scale at the correct relative pitch level. Thus the overlearned scale pattern, which is common to a large number of melodies, consists of a fixed pattern of pitch classes, and thus provides the sizes of the intervals required to realize the melody.

This contour + scale approach contrasts with the theory that melodies are represented as a combination of contour and pitch intervals. For example, Trainor, McDonald, and Alain (2002) say: "Melodic information is thought to be encoded in two different 'relative pitch' forms, a domain-general contour code (up/down pattern of pitch changes) and music-specific interval code (exact pitch distances between notes)." And Trehub and Hannon (2006) say: "Adults' ability to recognize or reproduce familiar tunes necessarily depends on their encoding of finer pitch relations, specifically, intervals, or precise pitch distances between successive tones." These statements appear to suggest a model in which the beginning of "Twinkle, Twinkle" would be encoded in terms of contour and intervals in semitones as (0, +7, 0, +2, 0, -2). In contrast, Dowling's (1978) theory suggests that it would be encoded by having the contour select the pitches (*do, do, sol, sol, la, la, sol*) out of an internalized scale in a "moveable *do*" system (where *do* is assigned to the pitch of the tonic in whatever key the melody is transposed to). There is considerable converging evidence in favor of the latter approach (see Dowling, 1991; Dowling, Kwak, & Andrews, 1995):

(1) Recognition of two-tone intervals isolated from melodic context and transposed to novel keys is relatively poor (Cuddy & Cohen, 1976; Lee, Janata, Frost, Martinez, & Granger, 2015).

(2) In contrast, transposition of familiar melodic patterns (e.g., the NBC chimes) is quite accurate, even for nonmusicians (Attneave & Olson, 1971). This suggests that the melodies are more fundamental, and that the intervals are retrieved in terms of the melodies, rather than vice versa. (In other words, the melody gives us a means of tapping the knowledge stored in the tonal scale systems.)

(3) The tonal hierarchy (Krumhansl, 1990) operates in terms of pitch classes, not intervals.

(4) And, in a closely related point, the dynamic tendencies of tones are defined in terms of pitch classes in the tonal hierarchy. The seventh scale degree tends upward toward the tonic whether it is 1 semitone away (in the major mode) or 2 (in the natural minor or the Dorian modes). The second scale degree tends downward from 2 semitones above the tonic (major) or 1 semitone (Phrygian mode).

(5) Harmonic intervals such as thirds and sixths remain similar when inverted, showing again that it is the pitch classes preserved by inversion that are important, and not the intervals between them (Balzano & Liesch, 1982).

(6) Familiar melodies can be recognized even when the note-to-note interval pattern has been destroyed by inserting interleaved distractor notes between the notes

of the melody (Dowling, 1973; Dowling, Lung, & Herrbold, 1987), or by scrambling the notes of the melody into several octaves while preserving the pitch classes (Dowling, 1984; Idson & Massaro, 1978).

(7) Dowling (1986) showed that listeners with moderate levels of musical training automatically encode novel melodies they hear in terms of scale steps, not intervals. In contrast, nonmusicians tend toward interval encoding, and musical professionals can use either strategy depending on task demands.

One of the consequences of the theory that melodies are represented as combinations of contours and scales is that wrong notes in the melodies that violate the tonal scale pattern should be much more obvious and easy to identify than wrong notes that violate expected interval sizes. It has already been observed that out-of-key notes "pop out" of an otherwise uniform stream of pitches (Janata, Birk, Tillmann, & Bharucha, 2003). The experiment reported here is the latest in a series of studies in which we directly compare the rapidity with which listeners detect out-of-key wrong notes versus interval-distorting wrong notes (see Dowling, 2008; 2009; Raman & Dowling, 2015). The main difference between this study and our previous studies is that here we include a condition involving unfamiliar melodies, as well as familiar melodies. We expected that the tendency to rely on key membership as an index of whether a note is a wrong note would be even stronger with unfamiliar melodies, since there are no hard and fast rules governing interval sizes in melodies.

## METHOD

*Participants*. Fifty-six students at the University of Texas at Dallas served in the experiment for partial course credit. Twenty-seven participants had less than 3 years of musical training, and 29 had 3 or more years and are characterized as moderately trained. Participants were assigned blindly to either the Familiar ($N$ = 21) or the Unfamiliar ($N$ = 35) condition.

*Stimuli.* The stimuli consisted of the first 16 to 24 notes of 32 familiar melodies that received the highest familiarity ratings in our previous studies (nursery tunes, folk tunes, patriotic songs, holiday songs, etc.), plus a stylistically similar group of unfamiliar folk songs drawn from Bronson (1976). These were all presented with their natural rhythms and at tempi that we judged to be comfortable for each melody. Each melody appeared twice in its respective session, and contained wrong notes in two different conditions. The wrong notes were introduced in a way that did not alter the contour of the melody. In the rare cases where the wrong note was part of a repeated pair of notes, the second note in the pair was also altered to match. There were eight kinds of wrong notes. The wrong notes were introduced by altering a target note either up or down from its original pitch, moving it 1 or 2 semitones, and landing on an in-key pitch or an out-of-key pitch. There were 64 trials in each session, in which these eight types of wrong note each occurred eight times. The wrong note could be introduced anywhere between the sixth note of the melody and the end. Previous research had shown that the up-down variable had negligible effects, and so we collapsed the data across that variable in the analyses reported here. The melodies were played by a MATLAB R2009b program as sequences of sine waves with linear on- and off-ramps to avoid clicks, and with 20-ms gaps between notes, and presented to listeners via high-quality headphones at comfortable levels. The program presented the stimuli in a different random order to each participant, and recorded their response times (RTs) to the wrong notes.

*Procedure.* Listeners were told that on each trial they would hear a melody that would often contain a wrong note, and that their task was to respond to the wrong note as quickly as possible. We told them to hold their fingers on the space bar and to press it as soon as they heard a wrong note. If they got to the end of the melody without hearing a wrong note, they were then to press the space bar to go on to the next trial. Each session lasted about 30 min.

## RESULTS

We analyzed the data separately in the unfamiliar and familiar conditions, in each case scoring the responses for hits (correct detections of a wrong note in a window of 300 to 3000 ms following the onset of an actual wrong note), and for median RTs for correct detections (hits) for each condition (collapsed across direction, up vs. down) for each participant.

*Unfamiliar Melodies*. Table 1 shows the number of hits (out of 16) for each condition for listeners at the two levels of musical training. The hit rates are quite low, as might be expected for unfamiliar tunes. We ran an Analysis of Variance (ANOVA) with musical training as a between-groups variable, and key membership (in vs. out) and distance from original note (1 vs. 2 semitones) as within-group variables. To our surprise, key membership was not an important factor in detection of the wrong notes. Only distance from the original pitch was significant, $F(1,33) = 5.87$, $p = .02$, with alterations of 1 semitone ($M = 2.26$) detected more often than alterations of 2 semitones ($M = 1.66$).

**Table 1. Number of hits with unfamiliar melodies (out of 16) by musically untrained participants and those with moderate amounts of training, for wrong notes that were in- or out-of-key, and 1 or 2 semitones removed from the original pitch. *N* = 35**

| Scale | Distance semitones | Untrained *N* = 18 | Moderate *N* = 17 | Means |
|---|---|---|---|---|
| IN | 1 | 2.39 | 2.24 | 2.31 |
| IN | 2 | 1.39 | 1.65 | 1.51 |
| OUT | 1 | 1.83 | 2.59 | 2.20 |
| OUT | 2 | 2.06 | 1.53 | 1.80 |
| | Means | 1.92 | 2.00 | |

Table 2 shows the means of the median RTs for each condition. Because the hit rates in Table 1 were so low, only 16 of the original 35 participants produced hits in all four conditions. Those 16 split 8 and 8 on musical training. In the ANOVA on their data, only key membership was significant, $F(1,14) = 14.22$, $p = .002$, with RTs to out-of-key wrong notes

(*M* = 1039 ms) almost twice as fast as RTs to in-key wrong notes (*M* = 1960 ms).

When we analyzed the hit rates of those listeners with complete RT records, we found no significant effects of any of the variables. That is, they did not show the effect of interval size alteration found with the larger group of listeners.

**Table 2. RTs (in ms) with unfamiliar melodies by musically untrained participants and those with moderate amounts of training, who achieved at least one hit in each of the four conditions, for wrong notes that were in- or out-of-key, and 1 or 2 semitones removed from the original pitch. *N* = 16**

| Scale | Distance semitones | Untrained *N* = 8 | Moderate *N* = 8 | Means |
|---|---|---|---|---|
| IN | 1 | 2237 | 1764 | 2000 |
| IN | 2 | 1993 | 1847 | 1920 |
| OUT | 1 | 1260 | 651 | 955 |
| OUT | 2 | 1281 | 964 | 1122 |
| | Means | 1693 | 1306 | |

*Familiar Melodies*. Table 3 shows the number of hits for each condition for listeners at the two levels of musical training. In the ANOVA, the effect of training was significant, $F(1,19) = 12.88$, $p = .002$, with moderately trained listeners scoring more hits ($M = 12.92$) than untrained listeners ($M = 9.50$). Key had a strong effect, with out-of-key wrong notes detected more often ($M = 12.62$ out of 16) than in-key ($M = 10.29$), $F(1,19) = 29.33$, $p < .001$. Distance had an effect also, $F(1,19) = 14.94$, $p = .001$, with 2-semitone alterations ($M = 12.12$) more noticeable than 1-semitone alterations ($M = 10.79$). And distance interacted with experience, $F(1,19) = 11.86$, $p = .003$, such that distance was more important for untrained listeners (a 2.88 item gain for greater distance) than for the moderately trained (a 0.16 item gain).

**Table 3. Number of hits with familiar melodies (out of 16) by musically untrained participants and those with moderate amounts of training, for wrong notes that were in- or out-of-key, and 1 or 2 semitones removed from the original pitch. *N* = 21**

| Scale | Distance semitones | Untrained *N* = 9 | Moderate *N* = 12 | Means |
|---|---|---|---|---|
| IN | 1 | 6.67 | 12.00 | 9.71 |
| IN | 2 | 9.56 | 11.83 | 10.86 |
| OUT | 1 | 9.44 | 13.67 | 11.86 |
| OUT | 2 | 12.33 | 14.17 | 13.38 |
| | Means | 9.50 | 12.92 | |

Table 4 shows the RTs for each condition for both groups of listeners. Moderately trained listeners responded more quickly ($M = 592$ ms) than untrained listeners ($M = 812$ ms), $F(1,19) = 11.74$, $p = .003$. RTs were shorter to out-of-key wrong notes (657 ms vs. 716 ms), $F(1,19) = 5.92$, $p = .03$, and to 2-semitone alterations ($M = 643$ ms) than to 1-semitone alterations ($M = 730$ ms), $F(1,19) = 8.65$, $p = .008$. And this distance effect appeared mainly in the untrained listeners (901 vs. 724 ms) than in the moderately trained (603 vs. 582 ms), $F(1,19) = 5.42$, $p = .03$.

The key membership X distance interaction approached significance, $F(1,19) = 3.02$, $p = .10$, in which the effect of key membership was stronger with a 2-semitone alteration ($M = 128$ ms) than with a 1-semitone alteration ($M = 46$ ms).

**Table 4. RTs (in ms) with familiar melodies by musically untrained participants and those with moderate amounts of training, for wrong notes that were in- or out-of-key, and 1 or 2 semitones removed from the original pitch. *N* = 21**

| Scale | Distance semitones | Untrained *N* = 9 | Moderate *N* = 12 | Means |
|---|---|---|---|---|
| IN | 1 | 932 | 594 | 739 |
| IN | 2 | 797 | 614 | 693 |
| OUT | 1 | 869 | 611 | 721 |
| OUT | 2 | 651 | 550 | 593 |
| | Means | 812 | 592 | |

## DISCUSSION

The most surprising result was that the listeners did not rely on key membership as a cue for wrong-note detection with the unfamiliar melodies. Distance of the wrong note from the original pitch had an effect, but it was the opposite of what was observed with the familiar melodies; that is, wrong notes 1 semitone away from the original were easier to detect than those 2 semitones away. We think that listeners were pursuing a strategy of noticing awkward sounding phrases, and taking those phrases as indicating the presence of wrong notes. Such a strategy risks producing RTs that are longer than even the 3-s window we provided. Hence it seems likely that many of the listeners who missed all the trials in one or another condition (and thus had to be deleted from the RT analysis) were following this strategy. When we looked at the detection data from the remaining listeners, it showed no effects of either distance or key membership.

These results complicate our understanding of the effects of experience on pitch encoding in memory for melodies. Dowling (1986), using an encoding specificity paradigm, demonstrated that moderately trained musicians automatically encode novel melodies they hear in terms of scale steps. However, if that is the case here, it seems that the moderately trained listeners are not developing a robust enough representation of the musical key of the unfamiliar melodies to use it in detecting wrong notes. Furthermore, it is clear from the present results that nonmusicians are not totally devoid of sensitivity to the scale steps (as they were in Dowling, 1986), since they show a strong effect of key membership on their RTs to the wrong notes in unfamiliar melodies.

Key membership is important in the detection of wrong notes in familiar melodies, leading to the detection of about 2.3 more wrong notes (out of 16) in the out-of-key conditions versus the in-key conditions. Distance was used as a cue, but with smaller effects, leading to an increase in detections of about 1.3 (out of 16). Note that the effect of distance was the opposite of that found with the unfamiliar melodies; here it was the wrong notes that were 2 semitones away from the original that were easier to detect. And distance was a more important cue for musically untrained listeners than for the moderately trained.

As with the unfamiliar melodies, RTs were strongly affected by key membership, with responses to out-of-key wrong notes about 59 ms faster. And there was a hint of an interaction between key membership and distance, such that both untrained and moderately trained listeners were especially quick to respond when an out-of-key wrong note was 2 semitones away from the original pitch. Distance affected RTs in general, but the effect was much more pronounced for the untrained listeners (about 177 ms faster for 2 semitones than for 1) than for the moderately trained (about 21 ms faster).

## CONCLUSION

The present results replicate earlier studies in finding strong effects of key membership on the detection of and RTs to wrong notes in familiar melodies. We also found strong effects of the distance in semitones between the wrong note and the correct pitch. In unfamiliar melodies, the picture was more complicated, in that lack of key membership facilitated fast RTs, but not detection. Distance between the wrong note and the correct pitch in unfamiliar melodies seemed to function in a more global way than in familiar melodies, perhaps as by producing awkward sounding phrases that suggested the occurrence of wrong notes.

## REFERENCES


Attneave, F., & Olson, R. K. (1971). Pitch as medium: A new approach to psychophysical scaling. *American Journal of Psychology, 84,* 147-166.

Balzano, G. J., & Liesch, B. W. (1982). The role of chroma and scalestep in the recognition of musical intervals out of context. *Psychomusicology, 2,* 331.

Bronson, B. H. (Ed.) (1976). *The Singing Tradition of Child's Popular Ballad*s. Princeton, NJ: Princeton University Press.

Dowling, W. J. (1973). The perception of interleaved melodies. *Cognitive Psychology, 5,* 322-337.

Dowling, W. J. (1978). Scale and contour: Two components of a theory of memory for melodies. *Psychological Review, 85,* 341-354.

Dowling, W. J. (1984). Musical experience and tonal scales in the recognition of octave-scrambled melodies. *Psychomusicology, 4,* 13-32.

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

Dowling, W. J. (1991). Pitch structure. In P. Howell, R. West, and I. Cross (Eds.) *Representing Musical Structure*. London: Academic Press, pp. 33-57.

Dowling, W. J. (2008, November). *Are melodies remembered as contour-plus-intervals or contour-plus-pitches?* Auditory Perception, Cognition, and Action Meetings (APCAM), Chicago, IL.

Dowling, W. J. (2009, November). *Directing auditory attention in terms of pitch, time, and tonal scale membership.* Poster session presented at the meeting of Psychonomic Society, Boston, MA.

Dowling, W. J., Kwak, S.-Y., & Andrews, M. W. (1995). The time course of recognition of novel melodies. *Perception & Psychophysics, 57,* 197-210.

Dowling, W. J., Lung, K. M.-T., & Herrbold, S. (1987). Aiming attention in pitch and time in the perception of interleaved melodies. *Perception & Psychophysics, 41,* 642-656.

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

Idson, W. L., & Massaro, D. W. (1978). A bidimensional model of pitch in the recognition of melodies. *Perception & Psychophysics, 24,* 551-565.

Krumhansl, C. L. (1990). *Cognitive Foundations of Musical Pitch*. New York: Oxford University Press.

Lee, Y.-S., Janata, P., Frost, C., Martinez, Z., & Granger, R. (2015). Melody recognition revisited: Influences of melodic Gestalt on the encoding of relational pitch information. *Psychological Bulletin & Review, 22,* 163-169.

Janata, P., Birk, J. L., Tillmann, B., & Bharucha, J. (2003). Online detection of tonal pop-out in modulating contexts. *Music Perception, 20,* 283-305.

Raman, R., & Dowling, W. J. (2015, August). *Effects of key membership and interval size in perceiving wrong notes: A cross-cultural study.* Paper session presented at the meeting of Society for Music Perception and Cognition (SMPC), Nashville, TN.

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,* 430-442.

Trehub, S. E., & Hannon, E. E. (2006). Infant music perception: Domain-general or domain-specific mechanisms? *Cognition, 100,* 73-99.