Fine-grained Implicit Memory for Key and Tempo

E. Glenn Schellenberg⊙, Michael W. Weiss, Chen Peng and Shayan Alam

Abstract

Listeners remember the pitch level (key) and tempo of musical recordings they have heard multiple times. They also have long-term implicit memory for the key and tempo of novel melodies heard for the first time in the laboratory. In previous research, however, the stimulus melodies were simple and repetitive and the changes in key or tempo were large. Here, we tested the limits of implicit memory for the key and tempo of more complex stimulus melodies. Musically trained and untrained listeners heard 12 novel melodies during an exposure phase and 24 (12 old, 12 new) during a subsequent test (recognition) phase. From exposure to test, half of the melodies were transposed up or down (changed in key) (Experiment 1), or sped up or slowed down (Experiment 2), but to varying degrees. Musically trained listeners displayed enhanced recognition, but transposing or changing the tempo of the melodies reduced performance similarly for all listeners. The effect of the key change did not wane as the transposition was reduced from 6 semitones to 1, but recognition in general was worse as the pitch range of the stimulus melodies increased. The magnitude of the tempo change had a very small effect on response patterns, but Bayesian analyses indicated that the observed data were more likely *without* considering magnitude. The results suggest that musically trained and untrained listeners have implicit memory for key and tempo that is remarkably fine-grained, even for melodies that are heard for the first time in the laboratory, such that small changes in either feature make a melody less recognizable.

Virtually everyone can recognize thousands of songs. In fact, "no one has even attempted to measure the limits of musical memory" (Halpern & Bartlett, 2010, p. 233) because the task is so daunting. Memory for melodies (tunes) remains good in old age, even among patients with Alzheimer's disease (Cuddy et al., 2012; Cuddy, Sikka, & Vanstone, 2015; Jacobsen et al., 2015; Vanstone et al., 2012). Nevertheless, when exposed to a new melody, listeners often remember the contour—the pattern of upward and downward movements in pitch—but not the exact intervals that define the tunes (Dowling & Fujitani, 1971; Edworthy, 1985). In the present investigation, we also examined memory for melody, asking questions that are fundamental to music perception and cognition. What is the nature of long-term mental representations for music? What features of a musical stimulus are remembered, and to what degree of accuracy?

Our focus was on long-term memory for the key (pitch level) and tempo (speed) of previously unfamiliar melodies. Melodies are useful as to-be-remembered stimuli for at least two reasons. One is that melodies in music (e.g., the violins in Beethoven's *Symphony No. 5*; the trumpet in *So What* by Miles Davis) are often instrumental and therefore free of the semantics of the lyrics. Another is that a melody can be recognized when it is presented in a novel key, at a novel tempo, and on an unusual instrument (a different

Department of Psychology, University of Toronto Mississauga, Mississauga, ON, Canada

Corresponding author:
Glenn Schellenberg, Department of Psychology, University of Toronto Mississauga, Mississauga, ON, Canada L5L 1C6.
Email: g.schellenberg@utoronto.ca

timbre; Schellenberg & Habashi, 2015). For example, it is relatively easy to imagine hearing *The Star-Spangled Banner* performed on a tuba, at a very low pitch, and at a very slow tempo. Why? Because a melody is mentally represented as relational information between consecutive notes in terms of their pitch and duration (Fujioka, Trainor, Ross, Kakigi, & Pantev, 2004; Trainor, McDonald, & Alain, 2002). One can therefore consider a melody to comprise abstract or schematic information, which defines its identity or meaning, and surface or veridical information, which is specific to a particular performance (e.g., Trainor, Wu, & Tsang, 2004).

Because a melody's identity is based on relational information, it is not surprising that most behavioral research has focused on memory for such relations. Indeed, memory for melodies relies on the ability to *ignore* irrelevant changes in surface features. Virtually everyone remembers the abstract structure of many melodies, including infants (Plantinga & Trainor, 2005). There is also evidence that relational information becomes more salient in melodic memory from 5 years to adulthood, with key becoming correspondingly less salient (Stalinski & Schellenberg, 2010).

For many years, the predominant view was that for mature listeners, surface features of music, particularly pitch, fade quickly from memory, except for individuals who have absolute pitch (AP)—the rare ability to name or produce a musical tone (e.g., Middle C) in isolation. For individuals without AP, memory for pitch was typically thought to last a maximum of 1 (Krumhansl, 2000) or 2 (Rakowski & Rogowski, 2007) minutes. Nevertheless, in the age of digital recordings, which have the exact same pitch, tempo, and timbre each time they are heard, it would seem odd if none of this information is retained in long-term memory.

In fact, we now know that long-term mental representations of real music contain much information about surface features. For example, when participants without AP are asked to sing songs with no canonical versions (e.g., *Yankee Doodle*) on different occasions, they do so with little variation in pitch height (Halpern, 1989). When asked to sing a familiar song from memory (e.g., *Let It Be*), their renditions are close to the pitch of the original recording (Frieler et al., 2013; Levitin, 1994). Although motor memory may play a role in these singing tasks, perceptual tasks provide converging results. For example, when adults and children are asked to identify whether a familiar recording is presented in the original key, performance is better than chance when foils are transposed (shifted upward or downward in key) by 2 semitones, and, in some instances, by only 1 semitone (Schellenberg & Trehub, 2003, 2008; Trehub, Schellenberg, & Nakata, 2008).

These findings are notable because 1 semitone is the smallest pitch change that separates adjacent notes in diatonic (major or minor) and chromatic scales. Even individuals with AP are often judged to be "correct" when they err

by 1 semitone on note-naming tasks (e.g., Vanzella & Schellenberg, 2010). Adult listeners without AP also exhibit better-than-chance memory for the pitch of the dial tone (Smith & Schmuckler, 2008) and the censor's bleep (Van Hedger, Heald, & Nusbaum, 2016). In the latter instance, approximately 74% and 83% of participants identified the true censor's bleep when the foil was mistuned by 1 and 2 semitones, respectively (chance = 50%). In a much more difficult task, listeners judged whether a single piano or violin tone (selected randomly from the equal-tempered chromatic scale) was in-tune or out-of-tune (i.e., mistuned by 0.5 semitones; Van Hedger, Heald, Huang, Rutstein, & Nusbaum, 2017). Performance was approximately 59% correct for non-AP musicians and 53% for nonmusicians, but still better than chance (50%) for both groups. Thus, in a world where the equal-tempered scale is fixed with standard tuning (A4 = 440 Hz), listeners exhibit memory for the pitch of tones they hear regularly.

Research on long-term memory for other surface features of familiar music is relatively sparse. In the case of tempo, production tasks show that listeners sing familiar recordings at a tempo that is close to the original (Levitin & Cook, 1996), and that the tempo of mothers' renditions of play-songs to their infants is almost identical across occasions (Bergeson & Trehub, 2002). In the case of timbre, listeners perform above chance levels when asked to identify recordings from 100-ms excerpts (Schellenberg, Iverson, & McKinnon, 1999). Such excerpts contain information about the overall timbre of the recordings, but no information about pitch or temporal relations between notes. In fact, after infants are exposed to a melody for 1 week, they show a novelty preference for a different melody, provided the familiar melody is performed on the same instrument at the same tempo (Trainor et al., 2004). In other words, timbre and tempo are incorporated into infants' mental representations of a familiarized melody.

But what do listeners remember after hearing a previously unfamiliar tune for the first time? The question is relevant because throughout most of human history (i.e., before the invention of the phonograph and standard tuning), every performance of every piece of music varied—at least to some degree—from one occasion to the next. Early research with non-AP participants focused on short-term memory for the pitch of pure tones, documenting that such memory was easily prey to interference. For example, when listeners make same-different judgments about standard and comparison tones, intervening tones impair performance, and more so than when spoken numbers are substituted for the words (Deutsch, 1970). Performance improves when one of the intervening tones is the same as the standard (Deutsch, 1975), or if all intervening tones are separated by small intervals, but declines when an intervening tone differs from the standard by a small amount (2/3 of a semitone; Deutsch, 1972), or if intervening tones are separated by large intervals (Deutsch, 1978). Other research on short-term memory for novel melodies has used

brief tone sequences and tasks that required same-different judgments (e.g., Bartlett & Dowling, 1980) or similarity ratings (e.g., Stalinski & Schellenberg, 2010).

In studies of long-term memory, however, questions about key and tempo make little sense to most listeners without music training. Thus, one methodological strategy has been to alter a surface feature (i.e., key, tempo, timbre) between initial exposure and test. If recognition suffers, listeners remembered the feature, at least implicitly. Research using this paradigm reveals that changing the timbre from exposure to test impairs recognition (Halpern & Müllensiefen, 2008; Peretz, Gaudreau, & Bonnel, 1998; Warker & Halpern, 2005), unless the change is to a very similar timbre (e.g., violin to viola, Lim & Goh, 2012). Changing articulation from *legato* to *staccato* can also cause decrements in recognition (Lim & Goh, 2013). Because a change in timbre impairs recognition even when the test phase is a full week after exposure (Schellenberg & Habashi, 2015), timbre may be a particularly memorable surface feature. Our focus here was on the accuracy of implicit memory for key and tempo, which, unlike timbre, vary continuously rather than categorically.

Other research with novel melodies confirms that a large change between exposure and test in terms of key (6 semitones) or tempo (64 beats per minute—bpm) impairs recognition after a short (e.g., 10 min: key—Cohen's $d = .566$, tempo—$d = .698$) and a long (1 day: key—$d = .431$, tempo—$d = .722$) delay, but not after 1 week (Schellenberg & Habashi, 2015). Effect sizes after 10 minutes or 1 day are far from trivial (i.e., at levels considered moderate to large). When melodies are changed in key *and* tempo, the decrement in recognition is additive rather than interactive (Schellenberg, Stalinski, & Marks, 2014). When the test phase occurs soon after exposure (same testing session), recognition suffers when tempo is changed by 15–20% (Halpern & Müllensiefen, 2008). Although adults have better long-term melodic memory than children, changing the key impairs recognition similarly for children and adults (Schellenberg, Poon, & Weiss, 2017), which raises the possibility that implicit memory for key is independent of individual differences in cognitive ability or exposure to music.

To the best of our knowledge, the present study is the first to examine the *accuracy* of long-term memory for the key and tempo of novel melodies by systematically varying the size of the transposition or tempo change between exposure and test. It is also unknown whether tempo changes smaller than 15–20% could impair melody recognition. In previous research, moreover, novel melodies were very short in duration (i.e., 4–10 s; Halpern & Müllensiefen, 2008; Kleinsmith & Neill, 2018), or longer but with a simple structure (AA'BA'—Schellenberg et al., 2014; Schellenberg & Habashi, 2015; AA'—Schellenberg et al., 2017), and much repetition within each melody. Thus, it is an open question whether previous findings would generalize broadly.

Our stimulus melodies were excerpted from Broadway showtunes. They were tuneful and tonal, but more complex than those used previously in terms of their pitch and temporal structures. We expected that increasing the size of the key change or tempo change from exposure to test would increase the detrimental impact on melody recognition, as one would expect from the Weber-Fechner law (i.e., larger stimulus change→larger perceived change). In line with this view, there is evidence that long-term memory for very brief melodies is disrupted more by large than by small transpositions (Kleinsmith & Neill, 2018). Nevertheless, other evidence of detailed memory for the pitch of familiar stimuli (Van Hedger et al., 2016, 2017) motivated us to predict that even a small key change would affect recognition for previously novel stimuli. For tempo, we expected that a 20% change would affect recognition (Halpern & Müllensiefen, 2008), but we had no reason to speculate that a 10% change might do so as well.

Finally, we expected that music training would be associated with explicit memory for melodies, as it is with explicit judgments in many other tests of musical ability (for review see Schellenberg & Weiss, 2013). Accounting for variance due to music training also allowed us to increase power to detect differences in recognition due to our stimulus manipulations. We doubted, however, that formal training in music would have an association with implicit memory for key or tempo because of the null developmental findings (Schellenberg et al., 2017), and because individual differences play a negligible role in other tests of implicit musical knowledge (Bigand & Poulin-Charronnat, 2006).

Experiment 1

Method

The research methods used in this report were approved by the Research Ethics Board of the University of Toronto, and conducted in accord with the principles outlined in the Declaration of Helsinki. All participants provided written informed consent.

Participants

The sample comprised 128 undergraduate students (29 males, 99 females, mean age 18.8 years, $SD = 2.1$) enrolled in an introductory psychology course, who were recruited without regard to music training and received partial course credit. None had hearing impairments (self-reports). The sample size (32 participants per condition) was identical to that of previous research, which reported reliable effects (Schellenberg et al., 2014, 2017; Schellenberg & Habashi, 2015). On average, participants had 4.7 years of music training ($SD = 5.9$), which was measured as the sum of private, group, and school-based lessons. As with previous samples of undergraduates from the same population (e.g.,

Table 1. Sources from which the stimuli were excerpted.

Song	Musical	Composer
Fish	La Revue des Ambassadeurs	Porter
High-Flyin' Wings on my Shoes	Les Girls	Porter
How'd You Like to Spoon with Me?	The Earl and the Girl	Kern
I'm in Love Again	Greenwich Village Follies of 1924	Porter
I'm Unlucky at Gambling	Fifty Million Frenchmen	Porter
I've a Shooting Box in Scotland	See America First	Porter
In the Shade of the New Apple Tree	Hooray for What!	Arlen
It Might as Well be Spring	State Fair	Rogers
It's a Lovely Day for a Walk	Blue Skies	Berlin
Last Night When We Were Young	Metropolitan	Arlen
Look for the Silver Lining	Sally	Kern
Napoleon	Jamaica	Arlen
Satan's Li'l Lamb	Americana	Arlen
She Didn't Say Yes	The Cat and the Fiddle	Kern
She was a Fair Young Maiden	The Pot of Gold	Porter
The Cocotte	Nymph Errant	Porter
The Sponge	Hitchy Koo of 1922	Porter
They Didn't Believe Me	The Girl from Utah	Kern
Till the Clouds Roll By	Oh Boy	Kern
When the Summer Moon Comes 'Long	Yale	Porter
Where Would You Get Your Coat?	Fifty Million Frenchmen	Porter
Which?	Paris	Porter
You're a Sentimental Guy	—	Berlin
You're the Prize Guy of Guys	Les Girls	Porter

Ladinig & Schellenberg, 2012; Schellenberg, Peretz, & Vieillard, 2008), the distribution of music training was skewed positively (mode = 0, median = 3). Thus, in the statistical analyses, music training was treated as a binary variable, with 70 moderately trained participants (> 2 years of lessons, hereafter *trained*) and 58 *untrained* participants (≤ 2 years).[1] Classifying unselected undergraduates in this manner is common in studies of memory for music (e.g., Bartlett & Dowling, 1980; Dowling, Kwak, & Andrews, 1995; Dowling & Tillmann, 2014; Schellenberg et al., 2014; Schellenberg & Habashi, 2015).

Materials

The stimuli were 24 piano melodies. Each was approximately 30 s in duration. The stimuli were excerpted from vocal melodies drawn from the *Great American Songbook*, a term used to describe songs composed for musical film or theatre (Broadway) between 1920 and 1960. Excerpts were selected from lesser-known works composed by Cole Porter, Jerome Kern, Irving Berlin, Howard Arlen, or Richard Rodgers, so that they would be tuneful and tonal, but unfamiliar to our participants. More detailed information is provided in Table 1. The excerpts were typically 16 measures, taken from the end of the first chorus. Although the melodies had many accidentals and changes in implied harmony, each melody ended on the tonic of the major key that corresponded to the key signature in the notation. For example, if the key signature had three flats, the melody ended on E-flat.

The stimulus melodies were originally entered note by note as MIDI (Musical Instrument Digital Interface) files using Finale NotePad 2010 (MakeMusic Inc.), which automatically added slight amplitude changes to reinforce the meter. GarageBand 5.1 (Apple Inc.) was then used to change the key and to save the stimuli as digital (MP3) sound files. Each melody was saved at the same tempo (142 bpm) but in five different keys, such that the median pitch (adjusted for duration) corresponded to a *standard* pitch (G4), or to higher pitches, specifically G#4, A4, A#4, or C#5 (1, 2, 3, or 6 semitones higher, respectively, than the standard). The tempo was the average of tempi used in previous research (Schellenberg et al., 2014; Schellenberg & Habashi, 2015). Customized software created with PsyScript (Slavin, 2007) was used to present stimuli over high-quality headphones and record responses.

Procedure

The procedure was identical to the key-change, between-subject condition of a melody-recognition task used previously (Schellenberg & Habashi, 2015; Schellenberg et al., 2014), with two exceptions: a) melodies were excerpted from the *Great American Songbook* instead of from British and Irish folksongs, and b) the key change varied in size (1, 2, 3, or 6 semitones). Participants were tested individually in a sound-attenuating booth and assigned randomly to one of four conditions, which differed only in the size of the key change.

Before the test session began, participants heard different versions of "Happy Birthday" to demonstrate that key is irrelevant to a tune's identity. All participants acknowledged readily that they understood the point. The actual test session had an exposure phase followed by a test phase. In the exposure phase, participants heard 12 of the melodies selected randomly from the set of 24. The 12 exposure melodies were presented in random order, followed by a second presentation in a different random order (no direct repetitions). To ensure that participants attended to each presentation of each melody, they were required to provide an emotionality rating that ranged from 1 (*very sad sounding*) to 6 (*very happy sounding*). These ratings were of no theoretical interest.

In all conditions, half of the melodies (6 of 12) were in the standard key, with a median pitch of G4 (G above middle C). The other half had a higher median pitch, which varied across conditions, specifically G#4 in the 1-semitone condition, A4 in the 2-semitone condition, A#4 in the 3-semitone condition, and C#5 in the 6-semitone condition. Melodies were equated for median instead of average pitch so that unusually high or low notes would not affect the overall pitch range, and because the median corresponded to an actual note in equal-tempered tuning. This method also ensured that within each condition, stimulus melodies varied in key. For example, a melody with a median pitch of G4 could be in the key of C, E-flat, F, G, or B-flat major.

After a delay of approximately 10 min, the test phase began. Participants were first instructed to ignore key changes and reminded that key is irrelevant to a melody's identity. They then heard all 24 stimulus melodies presented in random order, such that 12 were *old* (heard in the exposure phase) and 12 were *new*. They rated on a 6-point scale whether they heard the melody in the exposure phase (1 = *Definitely new*, 6 = *Definitely old*). Half of the new melodies were presented in the standard key. The other half were presented in a key 1, 2, 3, or 6 semitones higher, depending on condition. Half of the old melodies, high or low, were transposed down or up, respectively, depending on condition (i.e., by 1, 2, 3, or 6 semitones). The design ensured that overall pitch range was not a cue to whether any melody was old or new, or transposed.

Results and discussion

For ease of interpretation, descriptive statistics are reported as percentage-correct scores in Table 2, with ratings of 4–6 considered to represent "recognition" and ratings of 1–3 representing "no recognition." This approach, even with transformation to *d'* scores, discards much detail from the ratings and reduces statistical power, because ratings of 4, 5, and 6 (and 1, 2, and 3) are considered to be identical. Thus, in the statistical analyses that follow, recognition was calculated with Area Under the Curve (AUC) scores derived separately for each listener from the receiver

Table 2. Mean performance (and SDs) as percent correct scores.

Experiment 1—Key Condition	Original key	Transposed
1 semitone	83.7 (9.4)	82.5 (11.3)
2 semitone*	86.1 (9.8)	82.3 (9.7)
3 semitone	82.1 (11.0)	80.0 (10.7)
6 semitone*	77.1 (16.7)	74.3 (15.9)

Experiment 1a—Key Condition	Original key	Transposed
2 semitone (new)	79.2 (12.0)	78.3 (12.8)

Experiment 2—Tempo Condition	Original tempo	Changed
10%	82.5 (13.2)	81.6 (13.2)
20%	83.0 (11.9)	80.2 (14.5)
30%	79.5 (13.8)	77.6 (15.7)
60%	80.4 (12.3)	74.5 (12.8)

*Same data used in Experiment 1a.

operating characteristic curve (as in Dowling et al., 1995; Dowling, Tillmann, & Ayers, 2001; Schellenberg et al., 2014, Schellenberg & Habashi, 2015). AUC is considered to be a bias-free measure of recognition accuracy that considers the degree of overlap between ratings for old and new melodies. If all old melodies have higher ratings than all new melodies, recognition is considered perfect whether the listener has a conservative bias (e.g., all old melodies rated 3 or 4, all new melodies rated 1 or 2), a liberal bias (e.g., all old melodies rated 5 or 6, all new melodies rated 3 or 4), or no bias. A score of 1.0 corresponds to perfect recognition (i.e., no overlap between ratings for old and new melodies), whereas a score of 0.5 represents chance performance (i.e., ratings for old and new melodies are indistinguishable). For each listener, we calculated separate AUC scores for original-key and transposed melodies from 18 original ratings (6 old, 12 new; the same 12 new melodies were used in both calculations). One-sample *t*-tests confirmed that mean levels of performance were above chance levels in all instances, *p*s < .001. In short, listeners remembered the melodies.

We then tested our central questions: whether some melodies (i.e., those in the same key at exposure and test) were remembered better than others (i.e., those that were transposed from exposure to test), and whether the magnitude of the transposition affected recognition. A mixed-design Analysis of Variance (ANOVA) with one repeated measure (key change: original or transposed) and two between-subjects variables (music training, condition) confirmed three of our predictions: a) musically trained listeners ($M = .898$, $SD = .091$) had better overall memory compared to untrained listeners ($M = .861$, $SD = .098$), $F(1, 120) = 4.49$, $p = .036$, partial $\eta^2 = .036$ (Figure 1); b) changing key from exposure to test had a detrimental effect on recognition, $F(1, 120) = 16.09$, $p < .001$, partial $\eta^2 = .118$ (original key: $M = .900$, $SD = .102$; transposed:
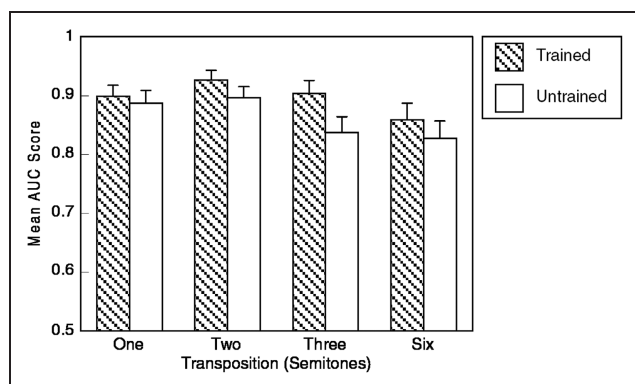
Figure 1. Mean recognition performance from Experiment 1 in each of the four testing conditions, reported separately for musically trained and untrained listeners. Error bars are standard errors of the mean. In general, trained listeners had better recognition, and recognition for all melodies deteriorated as the pitch variability in the stimulus set increased. These factors did not interact.
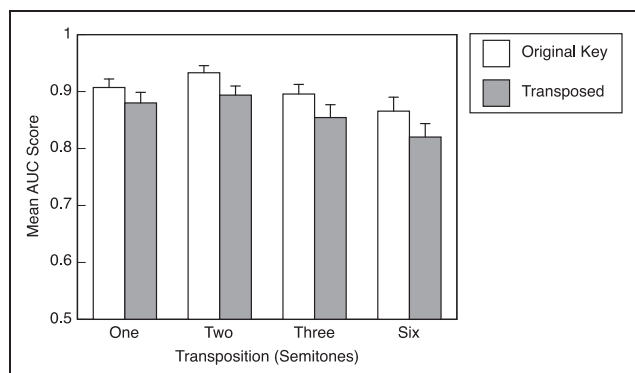

Figure 2. Mean recognition performance from Experiment 1 in each of the four testing conditions, reported separately for melodies that were re-presented in the original key or transposed. Error bars are standard errors of the mean. In general, original-key melodies were better recognized, but this effect was independent of the size of the transposition.

$M = .862$, $SD = .118$; Figure 2); and c) there was no two-way interaction between key change and music training, $p > .1$ (Figure 1). As shown in Figure 1, there was also an unexpected but relatively small main effect of condition, $F(3, 120) = 3.25$, $p = .024$, partial $\eta^2 = .075$, which arose because recognition for same-key *and* transposed melodies deteriorated linearly as the key change increased, $p = .011$. There was no three-way interaction, $p > .5$.

Contrary to one of our main predictions, the effect of the transposition was independent of its size, and not even close to statistical significance, $p > .9$ (i.e., no two-way interaction between key change and condition, Figure 2). In other words, even though the key change had a robust detrimental effect on recognition, the effect was similar whether the transposition was small (1 semitone) or large (6 semitones). Attempts to maximize power in this regard

(e.g., direct comparison of the smallest and largest transpositions, linear trend analysis) also led to null results.

Thus, two of the most important results stemmed from null findings, and a third finding was unexpected. Accordingly, we repeated the main analysis, which used Null Hypothesis Significance Testing (NHST), with Bayesian statistics conducted using JASP version 0.9.2 (JASP Team, 2019). Bayesian statistics help to determine whether observed effects were likely to stem from a lack of power (for null results) or a Type I error (for positive results).[2] The results were consistent with those from the traditional ANOVA, and confirmed that the best model of the observed data comprised main effects of key change, condition, and music training. In fact, the observed data were 596 times more likely (BF_{10}) under a model that included these three independent variables than they were with the null model. When we removed each independent variable from the main-effects model one at a time, the observed data were 1.67, 1.92, and 170 times *less* likely, respectively, when music training, condition, and the key change were excluded. Adding the interaction term between key change and condition provided a substantially *poorer* explanation of the observed data ($BF_{01} = 19.4$). Adding the interaction between key change and music training had a much smaller detrimental effect ($BF_{01} = 1.94$).

In sum, the Bayesian analysis provided strong evidence that the key change reduced recognition accuracy, and that the magnitude of the change had no effect. There was also relatively weak evidence for main effects of music training and condition, and for the lack of an interaction between key change and music training. When considered jointly with traditional NHST, it is clear that musically trained and untrained listeners exhibited implicit memory for key that was similar between groups, yet unexpectedly impervious to the size of the transposition.

Experiment 1a

In Experiment 1, recognition of original-key *and* transposed melodies decreased as the pitch range of the stimulus set became more heterogeneous, which was unexpected. One possibility is that as variation in pitch height of the stimulus melodies increased, listeners attended more to this variation, and, consequently, less to the relations that define the melodies. Nevertheless, as the pitch range of the stimulus set increased, so did the average pitch height of the stimulus melodies. To remedy this interpretive problem, we collected data from another sample of participants, who were tested in a new 2-semitone condition. In this new condition, the pitch range in the exposure phase was equivalent to that of the original 6-semitone condition, but the magnitude of the transposition was identical to that of the original 2-semitone condition. If a larger pitch range distracts from encoding the pitch relations that define the individual melodies, recognition performance should be equivalent to the original 6-semitone condition and inferior

to the original 2-semitone condition. The results from Experiment 1 also motivated us to predict that transposing a melody from exposure to test would be similarly detrimental to recognition in all instances, and independent of music training.

Method

Participants

The sample included data from 64 participants who were tested in Experiment 1: those in the 2-semitone and 6-semitone conditions. An additional 32 new participants were recruited as in Experiment 1 and tested in a new 2-semitone condition. The entire sample of 96 participants was similar demographically to Experiment 1 (71 women, 25 men; age: $M = 18.9$, $SD = 2.1$). On average, participants had 3.9 years of music training ($SD = 5.3$, mode = 0, median = 2). For statistical analysis, 47 were considered to be musically trained (> 2 years) and 49 untrained (≤ 2 years).

Materials

The stimulus melodies were the same as those from Experiment 1, but each was saved in four different keys: the standard key (median pitch G4), 2 semitones higher (median pitch A4), 4 semitones higher (median pitch B4), and 6 semitones higher (median pitch C#5).

Procedure

The procedure was the identical to Experiment 1 (2-semitone and 6-semitone conditions) except for the new 2-semitone condition. Its exposure phase was identical to the 6-semitone condition: 12 melodies were selected randomly from the set of 24, with 6 melodies presented in the standard key and 6 presented in the key 6 semitones higher. During the test phase, however, half of the old melodies that were originally presented in the standard key were transposed up 2 semitones, whereas half of the old melodies that were originally presented in the higher key were transposed down 2 semitones. The 12 new melodies were divided equally among the four pitch levels.

Results and discussion

We first confirmed that performance was better than chance levels in the new 2-semitone condition for both original-key and transposed melodies, ps < .001. Response patterns then were analyzed with a mixed-design ANOVA that had one repeated measure (key: original or transposed) and two between-subjects variables (condition, music training). Transposing the old melodies from exposure to test had a negative impact on recognition, as expected, $F(1, 90) = 9.06$, $p = .003$, partial $\eta^2 = .091$ (original key: $M = .888$, $SD = .115$; transposed: $M = .854$, $SD = .114$), but this

effect was similar across the three conditions (no two-way interaction), $F < 1$. No other interactions were significant, ps > .2. Other main effects confirmed that musically trained listeners ($M = .894$, $SD = .097$) had better recognition compared to their untrained counterparts ($M = .849$, $SD = .100$), $F(1, 90) = 4.17$, $p = .044$, partial $\eta^2 = .044$, and that recognition accuracy varied across conditions, $F(2, 90) = 4.30$, $p = .016$, partial $\eta^2 = .087$. Follow-up planned orthogonal contrasts were consistent with predictions. Performance was better in the original 2-semitone condition ($M = .913$, $SD = .071$) than it was in the other two conditions, $F(1, 90) = 7.79$, $p = .006$, partial $\eta^2 = .080$, which did not differ, $F < 1$ (6-semitone: $M = .843$, $SD = .116$; new 2-semitone: $M = .857$, $SD = .099$).

When we repeated the main analysis using Bayesian statistics, the best model of the observed data was consistent with NHST. It comprised the three main effects, and the observed data were 70.7 times more likely with this model compared to the null model. Removing each main effect from the model one at a time revealed that the observed data were 9.51, 1.66, and 3.75 times less likely, respectively, for key change, music training, and condition. The observed data were also 6.94 and 4.86 times less likely, respectively, when the interaction between condition and key change, or the interaction between music training and key change, was added to the model.

In sum, the findings from Experiment 1a were consistent with our proposed explanation of the results from Experiment 1. As the stimulus melodies became more heterogeneous in terms of pitch height during the exposure phase, subsequent recognition of the melodies declined. Transposing stimulus melodies from exposure to test also negatively affected recognition, but the magnitude of the transposition (6 vs 2 semitones) was irrelevant. Finally, music training had a small positive association with explicit recognition, but no association with the same-key memory advantage.

Experiment 2

In Experiment 2, we changed our focus, asking whether a change in *tempo* from exposure to test would negatively affect recognition, and whether a larger change in tempo would have a stronger detrimental impact.

Method

Participants

A new sample of 128 participants (91 women, 37 men, age: $M = 18.5$ years, $SD = 1.1$) was recruited as in Experiment 1. They had 4.5 years of music lesson on average ($SD = 5.5$, mode = 0, median = 3). For statistical analyses, 68 were considered to be musically trained (> 2 years) and 60 were untrained (≤ 2 years).

Materials

The stimulus melodies were those from Experiment 1, except they were always presented in the standard key (median pitch = G4). Tempo changes varied across conditions, however, in a manner that was designed to parallel the transpositions of Experiment 1. Previous results from three different samples of listeners documented that a change in tempo from 110 bpm to 174 bpm (or vice versa) was statistically equivalent in terms of psychological salience to a transposition of 6 semitones (Schellenberg et al., 2014; Schellenberg & Habashi, 2015). In fact, these manipulations were originally selected based on pilot testing designed specifically to determine changes that were equivalent (Schellenberg et al., 2014). Accordingly, stimulus melodies in the largest tempo-change condition were saved at these two tempi, which represented a difference of approximately 60% (174/110 = 1.582). To mirror the key-change manipulation of Experiment 1, other conditions had stimuli that differed in tempo by approximately 30%, 20%, or 10%, with mean tempo in all conditions fixed at 142 bpm, so that the testing session did not vary in overall duration, and was identical to Experiments 1 and 1a. In the 30% condition, the slow and fast tempi were 123 and 161 bpm, respectively, in the 20% condition they were 129 and 155 bpm, and in the 10% condition they were 135 and 149 bpm. (In GarageBand, tempo in bpm must be an integer.)

Procedure

The procedure was identical to Experiment 1 except that stimulus melodies varied in tempo rather than key, and the initial demonstration informed participants that tempo is irrelevant to a melody's identity. In each of four conditions (i.e., tempo change: 10%, 20%, 30% and 60%), melodies were presented at a slow and a fast tempo. In the exposure phase, 12 melodies from the set of 24 were selected randomly, with 6 assigned to the slow tempo and 6 to the fast tempo. In the test phase, 3 of the 6 old-slow melodies were sped up to the fast tempo, whereas 3 of the 6 old-fast melodies were slowed down to the slow tempo. The other 6 old melodies were identical at exposure and test. The 12 new melodies in the test phase were assigned randomly but equally to the slow and fast tempi. Thus, within any condition, tempo was not a cue to whether a melody was old or new, or to whether an old melody was changed in tempo from exposure to test.

Results and discussion

Descriptive statistics are reported as percentage-correct scores in Table 2. As in Experiment 1, ratings of 4–6 were considered to represent "recognition" and ratings of 1–3 "no recognition." For statistical analysis, two AUC scores were calculated for each listener from 18 original ratings (6 old, 12 new), one for original-tempo melodies and another
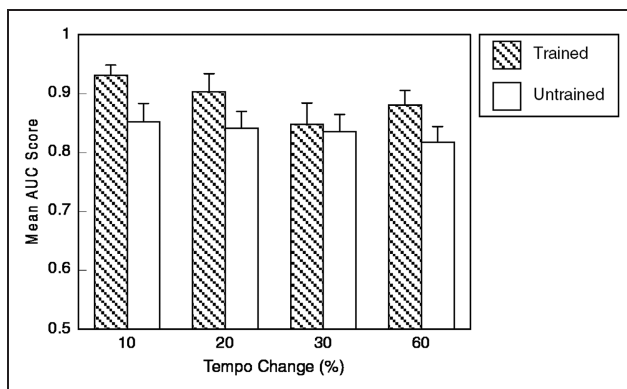

Figure 3. Mean recognition performance from Experiment 2 in each of the four testing conditions, reported separately for musically trained and untrained listeners. Error bars are standard errors of the mean. In general, trained listeners had better recognition, but the tempo change did not affect overall performance or interact with music training.
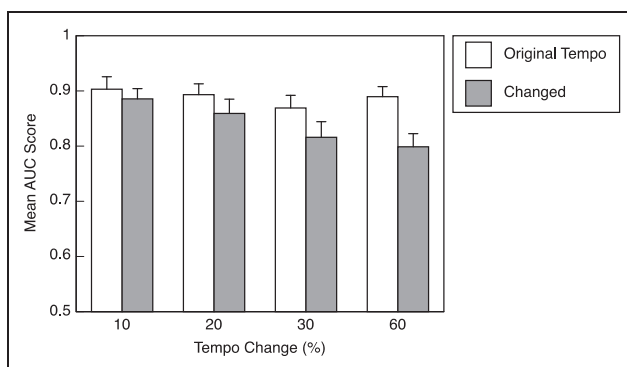

Figure 4. Mean recognition performance from Experiment 2 in each of the four testing conditions, reported separately for melodies that were re-presented at the original tempo or changed in tempo. Error bars are standard errors of the mean. In general, memory was better for original-tempo than for tempo-changed melodies. The advantage appeared to increase with the magnitude of the tempo change when NHST was used, but not with Bayesian analysis.

for changed-tempo melodies. Performance was above chance levels for both measures in all four conditions, $ps < .001$. A mixed-design ANOVA with one repeated measure (tempo: original or changed) and two between-subjects variables (music training, condition) revealed a main effect of music training, $F(1, 120) = 7.30$, $p = .008$, partial $\eta^2 = .057$. As shown in Figure 3, trained listeners ($M = .891$, $SD = .121$) had better recognition compared to untrained listeners ($M = .836$, $SD = .110$). As shown in Figure 4, a main effect of tempo confirmed that recognition was enhanced when melodies were presented at the same tempo during exposure and test (original tempo: $M = .889$, $SD = .118$; changed tempo: $M = .840$, $SD = .141$), $F(1, 120) = 27.78$, $p < .001$, partial $\eta^2 = .188$. This main effect was qualified, however, by a relatively small interaction between tempo and condition,

$F(3, 120) = 2.89$, $p = .038$, partial $\eta^2 = .067$. No other main effects or interactions were significant, ps $> .1$. The lack of a main effect of condition meant that, in contrast to the results from Experiments 1 and 1a, greater variance in tempo at encoding did *not* impair subsequent melody recognition.

The interaction between tempo and condition was investigated further by comparing recognition of same- and different-tempo melodies separately for each condition. As shown in Figure 4, as the tempo change became smaller, the detrimental impact on recognition was attenuated. Performance was better for same-tempo melodies in the 60% condition, $p < .001$, the 30% condition, $p = .008$, and the 20% condition, $p = .015$, but not in the 10% condition, $p > .1$ (one-tailed tests). Because there was no three-way interaction, this pattern was independent of music training.

Bayesian analyses raised doubts, however, about the reliability of the two-way interaction between tempo and condition that was observed with NHST. In fact, the probability of the observed data was highest for a model that comprised *only* main effects of the tempo change and music training ($BF_{10} = 931 \times 10^2$). Removing the main effect of music training and tempo change, respectively, reduced the probability of the observed data by factors of 6.30 and 153 $\times 10^2$. Adding the main effect of condition and the interaction between tempo and condition also *reduced* the likelihood of the observed data by a factor of 1.90 (BF_{01}), which provides support (albeit weakly) for the null hypothesis. Similarly, adding the interaction between music training and the tempo change reduced the likelihood of the observed data by a factor of 5.38 (BF_{01}), which provides moderate to strong support for the null hypothesis.

As in Experiment 1, then, the bulk of explained variance was due to a change in a surface feature, specifically tempo in the present experiment. Music training was also associated with better overall recognition, but it did not moderate the detrimental effect of the tempo change. Although the magnitude of the tempo change may have affected response patterns to some degree, the weak results from NHST ($.01 < p < .05$) combined with the results from the Bayesian analysis, suggest that if such an effect exists, it is small.

General discussion

We tested listeners' recognition of melodies that were either the same at test as exposure, shifted in key (transposed higher or lower), or changed in tempo (presented faster or slower). In general, melody recognition was well above chance levels whether the previously unfamiliar melodies were transposed or changed in tempo. Nevertheless, transposing melodies from exposure to test negatively affected recognition, as did changing tempo. For such effects to emerge, listeners must have retained information about key and tempo in their mental representations of the melodies, in addition to the pitch and temporal relations that defined the tunes. As predicted, effects of

the key and tempo change were independent of music training, even though musically trained individuals had moderately (but consistently) better explicit memory for the melodies. Finally, for both key and tempo, the magnitude of the change mattered little in terms of its negative impact on recognition. This last finding suggests that memory for key and tempo is finely tuned, such that a small change is similar to a large change in terms of its impact on recognition.

Our results replicated and extended results from previous studies that reported long-term memory for key (or pitch height) and tempo. For frequently heard tones, such as the dial tone (Smith & Schmuckler, 2008) or the censor's bleep (Van Hedger et al., 2016), listeners appear to remember their pitch, such that a 1-semitone deviation (i.e., approximately 6% deviation in frequency) is noticed at above-chance levels (see also Van Hedger et al., 2017). Listeners remember the key of familiar recordings with similar accuracy (e.g., Schellenberg & Trehub, 2003). For tempo, the results from Experiment 2—with decrements in recognition when tempo changes were 20% or greater—were in line with those of Halpern and Müllensiefen (2008), who reported that melody recognition suffered after a 15–20% change in tempo. The present findings generalized these previous results to more complex stimulus melodies that were heard for the first time in the laboratory.

In principle, if a larger sample were tested, performance could be affected significantly with a tempo change of 10% or smaller. We doubt, however, that increasing sample size to detect very small effects is the way forward for future research. With a sample of 128 participants, the interaction between tempo change and condition barely passed the threshold for statistical significance with NHST, and Bayesian analysis suggested that the observed data were more likely with a model that does *not* include this interaction. Perhaps a more sensitive outcome measure (e.g., mismatch negativity with EEG) would help to clarify the accuracy of implicit memory for musical tempo.

For key, the story was similar but even clearer. Transposing melodies from exposure to test negatively affected recognition, but the size of the transposition was irrelevant. More specifically, a) the interaction between the key change and its magnitude was not significant with NHST, b) Bayesian analyses provided strong support for the null hypothesis, and c) the decrement in recognition due to a 1-semitone transposition did not differ statistically from the decrement observed in any other condition. Although clear, this result seems counterintuitive. More generally, for key *and* for tempo changes, why were responses patterns inconsistent or only weakly consistent with the Weber-Fechner law, which is a well-established psychophysical principle?

For key, fundamental frequency is a continuous dimension that determines the perception of pitch, yet chromatic and diatonic scales subdivide frequency into 1- or 2-semitone bins (i.e., discontinuous categories) which are fixed with standard tuning (A4 = 440 Hz) across musical

genres and performances. Hence, the clearer results for key than for tempo may be attributable, at least in part, to some type of categorical perception that is relevant for key but not for tempo. This interpretation is supported by empirical evidence showing that musically trained and untrained listeners perform above chance levels at determining whether individual tones are in-tune or out-of-tune with the equal-tempered scale (Van Hedger et al., 2017). Through simple exposure, pitch categories may be learned, such that when a tone is presented midway in pitch between two "properly" tuned tones that are adjacent on the chromatic scale, listeners sense that something is not quite right. By contrast, tempo is almost completely continuous, except for the fact that it typically has an integer value for bpm.

Although this interpretation helps to explain why the results in Experiment 1 were somewhat cleaner than those in Experiment 2, it does not account for the major finding—that listeners incurred recognition deficits when a melody was changed in key *or* tempo, and that the size of the change mattered little. One possibility is that the phenomenon is a by-product of enhanced sensitivity to relatively small changes in the pitch or tempo of *speech*, which signal different speakers, and ultimately who is a friend or foe. Within-individual changes in pitch or tempo also signal an individual's mood, and the likelihood of a pleasant (safe) or unpleasant (harmful) interaction. We speculate that this perceptual sensitivity, which has obvious adaptive advantages, originates with speech but extends to music and perhaps to other auditory stimuli that vary in pitch or tempo. Pitch and tempo changes in speech *and* music are markers of communicative intent (e.g., emotion) *within* individuals, whereas average pitch height and tempo are reliable markers of differences *between* individual speakers, providing cues to a speaker's identity across contexts. Indeed, this view is plausible because specific changes in pitch or time are associated with the same emotions in both speech and music performance (Juslin & Luakka, 2003; Scherer, 1995). For example, anger is associated with fast tempo, whereas sadness is associated with low pitch, in both speech and music.

Changes in key or tempo also had *different* impacts on melody recognition, however, as evidenced by an incidental finding of Experiment 1, which was replicated in Experiment 1a, but did not extend to Experiment 2. As pitch-height variability increased in the stimulus set during the exposure phase, subsequent recognition of the melodies decreased, whether or not the melodies were transposed. Large inter-melody differences in average pitch height appeared to capture listeners' attention, which reduced their focus on, and subsequent memory for, pitch relations that defined the melodies. The issue of involuntary attention extends across domains, such that changes in an incidental dimension (e.g., a secondary task, a particular speaker) draw attention and reduce processing capacity for the principal dimension (e.g., the primary task, the content

of the utterance; e.g. Kahneman, 1973; Palmeri, Goldinger, & Pisoni, 1993).

Key changes but not tempo changes also change *spectral* information, which incorporates pitch (corresponding to the fundamental frequency) *and* timbre (corresponding to the dynamic and static structure of the harmonic series). Pitch and timbre are known to interact, such that timbre changes influence the perception of pitch, and pitch changes influence the perception of timbre (Krumhansl & Iverson, 1992; Melara & Marks, 1990; Moore & Glasberg, 1998). Timbre changes can also distort the perception of pitch relations (Russo & Thompson, 2005). In fact, the musical term *tessitura* refers to a portion of the entire pitch range of a particular voice, instrument, or piece: the portion that is the average or most common (Soanes & Stevenson, 2005), or the portion that has the most aesthetically pleasing timbre (Randel, 1986). Thus, key changes in the present experiment would have been accompanied by perceived changes in timbre, which provide important and memorable source cues in audition more generally (e.g., speech perception; Nygaard, Sommers, & Pisoni, 1994; Nygaard & Pisoni, 1998; Palmeri et al., 1993). In turn, these timbre cues may have provided an additional source of heterogeneity that served to capture listeners' attention. In previous research (Schellenberg & Habashi, 2015), a change in timbre impaired recognition to an equal extent regardless of the delay between exposure and test (10 min, 1 day, 1 week). In sum, human listeners may be excellent processors of spectral information (i.e., with relatively precise and enduring mental representations) that is conveyed by pitch in speech and key in music.

The present findings also inform accounts of AP and its ontogeny (Deutsch, 2013; Zatorre, 2003). When considered jointly with previous results, it is clear that the distinction between individuals with and without AP is more nuanced than common wisdom dictates, because: a) performance of AP possessors is far from being absolute, and b) mental representations of nonpossessors are more absolute than we once thought. In the former instance, *absolute* implies *context-free*, yet the context influences the performance of AP possessors (e.g., Vanzella & Schellenberg, 2010), such that exposure to a piece of music or tone sequence that is *flattened* (mistuned to be lower than equal temperament) leads to subsequent judgments of tones or tone sequences that are *sharpened* (Hedger, Heald, & Nusbaum, 2013; Van Hedger, Heald, Uddin, & Nusbaum, 2018). Moreover, when listeners are asked to identify a designated note presented in a sequence of tones, performance slows down as the foils vary more in pitch or timbre (Van Hedger, Heald, & Nusbaum, 2015). The results reported here replicate and extend other findings showing that individuals without AP or any formal music training recognize the key or pitch height of familiar auditory stimuli (Schellenberg & Habashi, 2015; Schellenberg & Trehub, 2003, Schellenberg et al., 2008, 2014, 2017; Smith & Schmuckler, 2008; Trehub et al., 2008; Van Hedger et al., 2016, 2017). Better

performance in this regard is also known to be correlated positively with general cognitive ability, specifically auditory working memory (Van Hedger, Heald, Koch, & Nusbaum, 2015; Van Hedger, Heald, & Nusbaum, 2018).

Across experiments, music training was associated positively with explicit recognition of the stimulus melodies, but it was independent of the decrement in performance caused by either a key or tempo change. The reliable effect in the former instance implies that the null effects in the latter were unlikely to stem from a lack of statistical power. Indeed, across Bayesian analyses, inclusion of the interaction between training and the tempo or key change *decreased* the likelihood of the observed data by factors of approximately 2 (Experiment 1) and 5 (Experiment 2). These response patterns fit nicely with a relatively large body of literature, which indicates that music training predicts performance on measures of explicit musical knowledge (Schellenberg & Weiss, 2013), but not on measures of implicit knowledge, for which performance tends to be relatively uniform across individuals (Bigand & Poulin-Charronnat, 2006). For example, musically trained individuals show enhanced performance on tasks that require them to identify *explicitly* whether a familiar melody is presented with an out-of-key note (Schellenberg & Moreno, 2010).

By contrast, tests of *implicit* knowledge of musical harmony reveal reliable effects that are independent of age and music training. When asked to identify whether a final chord of a chord sequence is sung with one syllable or another, or presented with one timbre or another, performance is enhanced for more stable chords in the musical context (i.e., tonic > subdominant; Bigand & Poulin-Charronnat, 2006). Moreover, such enhancements are similar for listeners who vary in formal music training (Bigand, Poulin, Tillmann, Madurell, & D'Adamo, 2003; Bigand, Tillmann, Poulin, D'Adamo, & Madurell, 2001), or in age and incidental exposure to music (Schellenberg, Bigand, Poulin-Charronnat, Garnier, & Stevens, 2005). These *harmonic-priming* effects are evident even in individuals with amusia, who have difficulty recognizing music or making other explicit judgments about music (Tillmann, Gosselin, Bigand, & Peretz, 2012; Tillmann, Peretz, Bigand, & Gosselin, 2007). Nevertheless, future research that included a group of active professional musicians could uncover differences in implicit memory for key and tempo that we were unable to detect.

Across analyses, musically trained participants had enhanced explicit recognition of the melodies in all instances when NHST was used, with Bayesian analyses confirming that the observed data were approximately two (Experiment 1) to six (Experiment 2) times more likely with the alternative hypothesis (i.e., groups differ) than with the null (i.e., no difference). What does this association tell us about the causal role of music training? Not much, because musical aptitude and music training are correlated (e.g., Law & Zentner, 2012; Wallentin, Nielsen, Friis-Olivarius, Vuust, & Vuust, 2010), and both variables co-vary with other individual differences, such as demographic background, cognitive ability, and personality (Corrigall, Schellenberg, & Misura, 2013; Swaminathan & Schellenberg, 2018). Low levels of natural musical ability (i.e., aptitude) are also accompanied by poor long-term memory for melodies (Nunes-Silva & Haase, 2012; Peretz, Champod, & Hyde, 2003). Moreover, genetic factors predict musical achievement and the propensity to practice music (Hambrick & Tucker-Drob, 2015). In short, formal music training may enhance memory for melodies, but musical aptitude—including memory for melodies—is also likely to increase the probability of taking music lessons.

In conclusion, the main findings of the present investigation highlight fine-grained implicit memory for key and tempo that is independent of formal training in music. In fact, recognition accuracy was impaired similarly by small or large changes in key or tempo. Many important questions remain unanswered about music and behavior, such as why music has the power to evoke emotions, inspire the imagination, and bring people together. Nevertheless, when one considers the precocious musical abilities of infant listeners (Trehub & Degé, 2016), the ubiquity of music across human cultures (Honing, 2018), and the present findings, it is clear that human listeners are exquisitely equipped to perceive and remember fundamental aspects of music, including fine-grained information about key and tempo.

Author note

Michael W. Weiss is now affiliated with the International Laboratory for Brain, Music, and Sound Research (BRAMS), Department of Psychology, University of Montreal, Montreal, Canada.

Author contribution

Declaration of conflicting interests

Funding

ORCID iD

E. Glenn Schellenberg https://orcid.org/0000-0003-3681-6020

Peer review

Notes

1. Throughout this report, results involving music training did not change when duration of training was square-root transformed (to reduce positive skew) and treated as a continuous variable in the statistical analyses (as in Swaminathan & Schellenberg, 2017, 2018; Swaminathan, Schellenberg, & Venkatesan, 2018). In other words, melodic memory in Experiments 1, 1a, and 2 improved as duration of training increased, but duration of training did not interact with any other variable or variables. These findings were confirmed with NHST and Bayesian statistics.

2. Results from Bayesian analyses are approximations based on sampling that vary slightly from one analysis to the next, particularly when models are complex. The analyses reported here involved comparing the observed data with 19 different models. Each Bayes factor reported in the manuscript is the average of 10 analyses. For ease of interpretation, we report BF_{10} for comparisons of the best model to the null model, and BF_{01} for comparisons of the best model to other models. Thus, all Bayes factors are greater than 1, and all are reported with 3-digit accuracy.

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*, 501–513. doi:10.1037/0096-1523.6.3.501.

Bergeson, T. R., & Trehub, S. E. (2002). Absolute pitch and tempo in mothers' songs to infants. *Psychological Science*, *13*, 72–75. doi:10.1111/1467-9280.00413.

Bigand, E., & Poulin-Charronnat, B. (2006). Are we "experienced listeners"? A review of the musical capacities that do not depend on formal musical training. *Cognition*, *100*, 100–130. doi:10.1016/j.cognition.2005.11.007.

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*, 159–171. doi:10.1037/0096-1523.29.1.159.

Bigand, E., Tillmann, B., Poulin, B., D'Adamo, D. A., & Madurell, F. (2001). The effect of harmonic context on phoneme monitoring in vocal music. *Cognition*, *81*, B11–B20. doi:10.1016/S0010-0277(01)00117-2.

Corrigall, K. A., Schellenberg, E. G., & Misura, N. M. (2013). Music training, cognition, and personality. *Frontiers in Psychology*, *4*, 222. doi:10.3389/fpsyg.2013.00222.

Cuddy, L. L., Duffin, J. M., Gill, S. S., Brown, C. L., Sikka, R., & Vanstone, A. D. (2012). Memory for melodies and lyrics in Alzheimer's disease. *Music Perception*, *29*, 479–491. doi:10.1525/mp.2012.29.5.479.

Cuddy, L. L., Sikka, R., & Vanstone, A. (2015). Preservation of musical memory and engagement in healthy aging and Alzheimer's disease. *Annals of the New York Academy of Sciences*, *1337*, 223–231. doi:10.1111/nyas.12617.

Deutsch, D. (1970). Tones and numbers: Specificity of interference in immediate memory. *Science*, *168*, 1604–1605. doi:10.1126/science.168.3939.1604.

Deutsch, D. (1972). Mapping of interactions in the pitch memory store. *Science*, *175*, 1020–1022. doi:10.1126/science.175.4025.1020.

Deutsch, D. (1975). Facilitation by repetition in recognition memory for tonal pitch. *Memory & Cognition*, *3*, 263–266. doi:10.3758/BF03212909.

Deutsch, D. (1978). Delayed pitch comparisons and the principle of proximity. *Perception & Psychophysics*, *23*, 227–230. doi:10.3758/BF03204130.

Deutsch, D. (2013). Absolute pitch. In D. Deutsch (Ed.), *The psychology of music (3rd ed.*, pp. 141–182). Amsterdam, the Netherlands: Elsevier. doi:10.1016/B978-0-12-381460-9.00005-5.

Dowling, W. J., & Fujitani, D. S. (1971). Contour, interval, and pitch recognition in memory for melodies. *Journal of the Acoustical Society of America*, *49*, 524–531. doi:10.1121/1.1912382.

Dowling, W. J., Kwak, S., & Andrews, M. W. (1995). The time course of recognition of novel melodies. *Perception & Psychophysics*, *57*, 136–149. doi:10.3758/BF03206500.

Dowling, W. J., & Tillmann, B. (2014). Memory improvement while hearing music: Effects of structural continuity on feature binding. *Music Perception*, *32*, 11–32. doi:10.1525/MP.2014.32.1.11.

Dowling, W. J., Tillmann, B., & Ayers, D. F. (2001). Memory and the experience of hearing music. *Music Perception*, *19*, 249–276. doi:10.1525/mp.2001.19.2.249.

Edworthy, J. (1985). Interval and contour in melody processing. *Music Perception*, *2*, 375–388. doi:10.2307/40285305.

Frieler, K., Fischinger, T., Schlemmer, K., Lothwesen, K., Jakubowski, K., & Müllensiefen, D. (2013). Absolute memory for pitch: A comparative replication of Levitin's 1994 study in six European labs. *Musicae Scientiae*, *17*, 334–349. doi:10.1177/1029864913493802.

Fujioka, T., Trainor, L. J., Ross, B., Kakigi, R., & Pantev, C. (2004). A comparison of automatic melodic contour and interval encoding in musicians and nonmusicians. *Journal of Cognitive Neuroscience*, *16*, 1010–1021. doi:10.1162/0898929041502706.

Halpern, A. R. (1989). Memory for the absolute pitch of familiar songs. *Memory & Cognition*, *17*, 572–581. doi:10.3758/BF03197080.

Halpern, A. R., & Bartlett, J. C. (2010). Memory for melodies. In M. R. Jones, R. R. Fay, & A. N. Popper (Eds.), *Springer handbook of auditory research: Vol. 36. Music perception* (pp. 233–258). New York, NY: Springer.

Halpern, A. R., & Müllensiefen, D. (2008). Effects of timbre and tempo change on memory for music. *Quarterly Journal of Experimental Psychology*, *61*, 1371–1384. doi:10.1080/17470210701508038.

Hambrick, D. Z., & Tucker-Drob, E. M. (2015). The genetics of music accomplishment: Evidence for gene-environment correlation and interaction. *Psychonomic Bulletin & Review*, *22*, 112–120. doi:10.3758/s13423-014-0671-9.

Hedger, S. C., Heald, S. L. M., & Nusbaum, H. C. (2013). Absolute pitch may not be so absolute. *Psychological Science*, *24*, 1496–1502. doi:10.1177/0956797612473310.

Honing, H. (Ed.). (2018). *The origins of musicality*. Cambridge, MA: MIT Press.

Jacobsen, J.-H., Stelzer, J., Fritz, T. H., Chételat, G., La Joie, R., & Turner, R. (2015). Why musical memory can be preserved in advanced Alzheimer's disease. *Brain*, *138*, 2438–2450. doi: 10.1093/brain/awv135.

JASP Team. (2019). *JASP* (Version 0.9.2) [Computer software].

Juslin, P. N., & Laukka, P. (2003). Communication of emotions in vocal expression and music performance: Different channels, same code? *Psychological Bulletin*, *129*, 770–814. doi:10.1037/0033-2909.129.5.770.

Kahneman, D. (1973). *Attention and effort*. Englewood Cliffs, NJ: Prentice-Hall.

Kleinsmith, A. L., & Neill, W. T. (2018). Recognition of transposed melodies: Effects of pitch distance and harmonic distance. *Psychonomic Bulletin & Review*, *25*, 1855–1860. doi:10.3758/s13423-017-1406-5.

Krumhansl, C. L. (2000). Rhythm and pitch in music cognition. *Psychological Bulletin*, *126*, 159–179. doi:10.1037/0033-2909.126.1.159.

Krumhansl, C. L., & Iverson, P. (1992). Perceptual interactions between musical pitch and timbre. *Journal of Experimental Psychology: Human Perception and Performance*, *18*, 739–751. doi:10.1037/0096-1523.18.3.739.

Ladinig, O., & Schellenberg, E. G. (2012). Liking unfamiliar music: Effects of felt emotion and individual differences. *Psychology of Aesthetics, Creativity, and the Arts*, *6*, 146–154. doi:10.1037/a0024671.

Law, L. N., & Zentner, M. (2012). Assessing musical abilities objectively: Construction and validation of the profile of music perception skills. *PLoS One*, *7*, e52508. doi:10.1371/journal.pone.0052508.

Levitin, D. J. (1994). Absolute memory for musical pitch: Evidence from the production of learned melodies. *Perception & Psychophysics*, *56*, 414–423. doi:10.3758/BF03206733.

Levitin, D. J., & Cook, P. R. (1996). Memory for musical tempo: Additional evidence that auditory memory is absolute. *Perception & Psychophysics*, *58*, 927–935. doi:10.3758/BF03205494.

Lim, S. W. H., & Goh, W. D. (2012). Variability and recognition memory: Are there analogous indexical effects in music and speech? *Journal of Cognitive Psychology*, *24*, 602–616. doi:10.1080/20445911. 2012.674029.

Lim, S. W. H., & Goh, W. D. (2013). Articulation effects in melody recognition memory. *Quarterly Journal of Experimental Psychology*, *66*, 1774–1792. doi:10.1080/17470218.2013.766758.

Melara, R. D., & Marks, L. E. (1990). Interaction among auditory dimensions: Timbre, pitch, and loudness. *Perception & Psychophysics*, *48*, 169–178. doi:10.3758/BF03207084.

Moore, B. C. J., & Glasberg, B. R. (1998). Frequency discrimination of complex tones with overlapping and non-overlapping harmonics. *Journal of the Acoustical Society of America*, *87*, 2163–2177.

Nunes-Silva, M., & Haase, V. G. (2012). Montreal battery of evaluation of Amusia: Validity evidence and norms for adolescents in Belo Horizonte, Minas Gerais, Brazil. *Dementia & Neuropsychologia*, *6*, 244–252. doi:10.1590/S1980-57642012DN06040008.

Nygaard, L. C., & Pisoni, D. B. (1998). Talker-specific learning in speech perception. *Perception & Psychophysics*, *60*, 355–376. doi:10.3758/BF03206860.

Nygaard, L. C., Sommers, M. S., & Pisoni, D. B. (1994). Speech perception as a talker-contingent process. *Psychological Science*, *5*, 42–46. doi:10.1111/j.1467-9280.1994.tb00612.x.

Palmeri, T. J., Goldinger, S. D., & Pisoni, D. B. (1993). Episodic encoding of voice attributes and recognition memory for spoken words. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*, 309–328. doi:10.1037/0278-7393.19.2.309.

Peretz, I., Champod, A. S., & Hyde, K. (2003). Varieties of musical disorders. *The Montreal battery of evaluation of Amusia. Annals of the New York Academy of Sciences*, *999*, 58–75. doi:10.1196/annals.1284.006.

Peretz, I., Gaudreau, D., & Bonnel, A.-M. (1998). Exposure effects on music preference and recognition. *Memory & Cognition*, *26*, 884–902. doi:10.3758/BF03201171.

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

Rakowski, A., & Rogowski, P. (2007). Experiments on long-term and short-term memory for pitch in musicians. *Archives of Acoustics*, *32*, 815–826.

Randel, D. M. (Ed.). (1986). *The new Harvard dictionary of music*. Cambridge, MA: Belknap.

Russo, F. A., & Thompson, W. F. (2005). An interval size illusion: The influence of timbre on the perceived size of melodic intervals. *Perception & Psychophysics*, *67*, 559–568. doi:10.3758/BF03193514.

Schellenberg, E. G., Bigand, E., Poulin, B., Garnier, C., & Stevens, C. (2005). Children's implicit knowledge of harmony in Western music. *Developmental Science*, *8*, 551–566. doi:10.1111/j.1467-7687.2005.00447.x.

Schellenberg, E. G., & Habashi, P. (2015). Remembering the tune and timbre, forgetting the key and tempo. *Memory & Cognition*, *43*, 1021–1031. doi:10.3758/s13421-015-0519 -1.

Schellenberg, E. G., Iverson, P., & McKinnon, M. C. (1999). Name that tune: Identifying popular recordings from brief excerpts. *Psychonomic Bulletin & Review*, *6*, 641–646. doi:10.3758/BF03212973.

Schellenberg, E. G., & Moreno, S. (2010). Music lessons, pitch processing, and g. *Psychology of Music*, *38*, 209–221. doi:10.1177/0305735609339473.

Schellenberg, E. G., Peretz, I., & Vieillard, S. (2008). Liking for happy and sad sounding music: Effects of exposure. *Cognition and Emotion*, *22*, 218–237. doi:10.1080/02699930701350753.

Schellenberg, E. G., Poon, J., & Weiss, M. W. (2017). Memory for melody and key in childhood. *PLoS One*, *12*, e0187115. doi:10.1371/journal.pone.0187115.

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:10.1007/s00426-013-0483-y.

Schellenberg, E. G., & Trehub, S. E. (2003). Good pitch memory is widespread. *Psychological Science*, *14*, 262–266. doi:10.1111/1467-9280.03432.

Schellenberg, E. G., & Trehub, S. E. (2008). Is there an Asian advantage for pitch memory? *Music Perception*, *25*, 241–252. doi:10.1525/mp.2008.25.3.241.

Schellenberg, E. G., & Weiss, M. W. (2013). Music and cognitive abilities. In D. Deutsch (Ed.), *The psychology of music (3rd ed.*, pp. 499–550). Amsterdam, the Netherlands: Elsevier. doi:10.1016/B978-0-12-381460-9.00012-2.

Scherer, K. R. (1995). Expression of emotion in voice and music. *Journal of Voice*, *9*, 235–248. doi:10.1016/S0892-1997(05)80231-0.

Slavin, S. (2007). *PsyScript* (Version 2.1) [software]. Retrieved from http://www.psych.lancs.ac.uk/software/psyScript.html.

Smith, N. A., & Schmuckler, M. A. (2008). Dial A440 for absolute pitch: Absolute pitch memory by non-absolute pitch possessors. *Journal of the Acoustical Society of America*, *123*, EL77–EL84. doi:10.1121/1.2896106.

Soanes, C., & A. Stevenson (Eds.). (2005). *Oxford dictionary of English* (2nd ed., Revised). Oxford, UK: Oxford University Press.

Stalinski, S. M., & Schellenberg, E. G. (2010). Shifting perceptions: Developmental changes in judgments of melodic similarity. *Developmental Psychology*, *46*, 1799–1803. doi:10.1037/a0020658.

Swaminathan, S., & Schellenberg, E. G. (2017). Musical competence and phoneme perception in a foreign language. *Psychonomic Bulletin & Review*, *24*, 1929–1934. doi:10.3758/s13423-017-1244-5.

Swaminathan, S., & Schellenberg, E. G. (2018). Musical competence is predicted by music training, cognitive abilities, and personality. *Scientific Reports*, *8*, 9223. doi:10.1038/s41598-018-27571-2.

Swaminathan, S., Schellenberg, E. G., & Venkatesan, K. (2018). Explaining the association between music training and reading in adults. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *44*, 992–999. doi:10.1037/xlm0000493.

Tillmann, B., Gosselin, N., Bigand, E., & Peretz, I. (2012). Priming paradigm reveals harmonic structure processing in congenital amusia. *Cortex*, *48*, 1073–1078. doi:10.1016/j.cortex.2012.01.001.

Tillmann, B., Peretz, I., Bigand, E., & Gosselin, N. (2007). Harmonic priming in an amusic patient: The power of implicit tasks. *Cognitive Neuropsychology*, *24*, 603–622. doi:10.1080/02643290701609527.

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. doi:10.1162/089892902317361949.

Trainor, L. J., Wu, L., & Tsang, C. D. (2004). Long-term memory for music: Infants remember tempo and timbre. *Developmental Science*, *7*, 289–296. doi:10.1111/j.1467-7687.2004.00348.x.

Trehub, S. E., & Degé, F. (2016). Reflections on infants as musical connoisseurs. In G. McPherson (Ed.), *The child as musician* (2nd ed., pp. 31–51). Oxford, UK: Oxford University Press.

Trehub, S. E., Schellenberg, E. G., & Nakata, T. (2008). Cross-cultural perspectives on pitch memory. *Journal of Experimental Child Psychology*, *100*, 40–52. doi:10.1016/j.jecp.2008.01.007.

Van Hedger, S. C., Heald, S. L. M., Huang, A., Rutstein, B., & Nusbaum, H. C. (2017). Telling in-tune from out-of-tune: Widespread evidence for implicit absolute intonation. *Psychonomic Bulletin & Review*, *24*, 481–488. doi:10.3758/s13423-016-1099 -1.

Van Hedger, S. C., Heald, S. L. M., Koch, R., & Nusbaum, H. C. (2015). Auditory working memory predicts individual differences in absolute pitch learning. *Cognition*, *140*, 95–110. doi:10.1016/j.cognition.2015.03.012.

Van Hedger, S. C., Heald, S. L. M., & Nusbaum, H. C. (2015). The effects of acoustic variability on absolute pitch categorization: Evidence for contextual tuning. *Journal of the Acoustical Society of America*, *138*, 436–446. doi:10.1121/1.4922952.

Van Hedger, S. C., Heald, S. L. M., & Nusbaum, H. C. (2016). What the [bleep?]: Enhanced absolute pitch memory for a 1000 Hz sine tone. *Cognition*, *154*, 139–150. doi:10.1016/j.cognition.2016.06.001.

Van Hedger, S. C., Heald, S. L. M., & Nusbaum, H. C. (2018). Long-term pitch memory for music recordings is related to auditory working memory precision. *The Quarterly Journal of Experimental Psychology*, *71*, 879–891. doi:10.1080/17470218.2017.1307427.

Van Hedger, S. C., Heald, S. L. M., Uddin, S., & Nusbaum, H. C. (2018). A note by any other name: Intonation context rapidly changes absolute note judgments. *Journal of Experimental Psychology: Human Perception & Performance*, *44*, 1268–1282. doi:10.1037/xhp0000536.

Vanstone, A. D., Sikka, R., Tangness, L., Sham, R., Garcia, A., & Cuddy, L. L. (2012). Episodic and semantic memory for melodies in Alzheimer's disease. *Music Perception*, *29*, 501–507. doi:10.1525/mp.2012.29.5.501.

Vanzella, P., & Schellenberg, E. G. (2010). Absolute pitch: Effects of timbre on note-naming ability. *PLoS One*, *5*, e15449. doi:10.1371/journal.pone.0015449.

Wallentin, M., Nielsen, A. H., Friis-Olivarius, M., Vuust, C., & Vuust, P. (2010). The musical ear test: A new reliable test for measuring musical competence. *Learning and Individual Differences*, *20*, 188–196. doi:10.1016/j.lindif.2010.02.004.

Warker, J. A., & Halpern, A. R. (2005). Musical stem completion: Humming that note. *American Journal of Psychology*, *118*, 567–585.

Zatorre, R. J. (2003). Absolute pitch: A model for understanding the influence of genes and development on neural and cognitive function. *Nature Neuroscience*, *6*, 692–695. doi:10.1038/nn1085.