

Remembering the melody and timbre, forgetting the key and tempo

E. Glenn Schellenberg¹ · Peter Habashi¹

Published online: 24 March 2015
© Psychonomic Society, Inc. 2015

Abstract The identity of a melody is independent of surface features such as key (pitch level), tempo (speed), and timbre (musical instrument). We examined the duration of memory for melodies (tunes) and whether such memory is affected by changes in key, tempo, or timbre. After listening to previously unfamiliar melodies twice, participants provided recognition ratings for the same (old) melodies as well as for an equal number of new melodies. The delay between initial exposure and test was 10 min, 1 day, or 1 week. In Experiment 1, half of the old melodies were transposed by six semitones or shifted in tempo by 64 beats per minute. In Experiment 2, half of the old melodies were changed in timbre (piano to saxophone, or vice versa). In both experiments, listeners remembered the melodies, and there was no forgetting over the course of a week. Changing the key or tempo from exposure to test had a detrimental impact on recognition after 10 min and 1 day, but not after 1 week. Changing the timbre affected recognition negatively after all three delays. Mental representations of unfamiliar melodies appear to be consolidated after only two presentations. These representations include surface information unrelated to a melody's identity, although information about key and tempo fades at a faster rate than information about timbre.

Keywords Memory · Music · Pitch · Tempo · Timbre · Melody · Absolute pitch

Music is an abstract domain, in the sense that a melody is recognized on the basis of relations in pitch and time between consecutive notes. As such, almost everyone can identify “Happy Birthday” or “Hey Jude” when the tune is played on a novel instrument (e.g., xylophone), in a novel key (pitch level), or at a novel tempo (speed). Although each of these surface transformations changes *all* of the tones, none causes recognition problems for the average listener if the pitch and time relations remain perceptible. One might speculate, then, that long-term mental representations of melodies are composed solely of relational information (Raffman, 1993). Indeed, in the case of pitch, the longstanding view is that only listeners with *absolute pitch* remember key after a delay lasting a minute or longer (Krumhansl, 2000). Absolute pitch is the rare ability to identify or produce a musical tone (e.g., middle C) in isolation (for a review, see Deutsch, 2013).

We now know, however, that musically untrained adults remember the key of a familiar recording that they have heard multiple times, just as they remember its tempo and timbre (i.e., the specific musical instrument). For example, they sing a melody from a familiar recording at close to the original key (Frieler et al., 2013; Levitin, 1994). When asked to judge whether a familiar recording is presented in the original key, performance is approximately 70 % or 60 % correct when the comparison is shifted in key (transposed) by two semitones or one semitone, respectively, exceeding chance levels (50 %) in both cases (Schellenberg & Trehub, 2003). Similar evidence has been found for adults' memory for the pitch of the dial tone (Smith & Schmuckler, 2008) and for children's (Schellenberg & Trehub, 2008; Trehub,

Electronic supplementary material The online version of this article (doi:10.3758/s13421-015-0519-1) contains supplementary material, which is available to authorized users.

✉ E. Glenn Schellenberg
g.schellenberg@utoronto.ca

¹ Department of Psychology, University of Toronto Mississauga, Mississauga, Ontario, Canada L5L 1C6

Schellenberg, & Nakata, 2008) and infants' (Volkova, Trehub, & Schellenberg, 2006) memory for the key of familiar recordings. Infants' memory for key is evident with lullabies sung expressively in a foreign language (Volkova et al., 2006), but not with computer-generated tunes presented in a piano timbre (Plantinga & Trainor, 2005).

The average undergraduate also sings the melody from a familiar recording at close to the original tempo ($\pm 4\%$; Levitin & Cook, 1996). In fact, their renditions often differ from the original by less than the just noticeable difference (JND) in tempo discrimination. Listeners can also recognize a 100-ms excerpt from a familiar recording at above-chance levels, but not if the excerpts are played backward (Schellenberg, Iverson, & McKinnon, 1999). Because such brief excerpts have no information about pitch and time relations, and backward presentation changes timbre but not key, the results implicate memory for the timbre of the recordings (i.e., overall sound quality, in this case). Better recognition of music when the timbre at test (piano or orchestral) matches the timbre at exposure provides corroborating evidence that listeners remember timbre, and that timbre may be an integral part of the identity of a musical piece (Poulin-Charronnat et al., 2004). Even infants exhibit memory for the timbre and tempo of a melody after 1 week of daily exposure (Trainor, Wu, & Tsang, 2004).

Hearing music identically (re: key, tempo, and timbre) across repetitions is a consequence of a relatively recent cultural artifact—reproductions in general, and digital recordings in particular. Throughout most of human history, the same tune typically changed from one presentation to the next, just as any live rendition of “Happy Birthday” differs from previous renditions. In live performances of music with no fixed-pitch instruments such as the piano or organ, key and tempo would change slightly from one rendition to the next until the advent of tuning forks in 1711 and metronomes in 1812 (Randel, 1986). Moreover, formal standardization of tuning, in which the A above middle C is set to 440 Hz, was not specified by the American Standards Association until 1936. Thus, it is important to ask whether listeners remember the surface features of previously unfamiliar melodies or of familiar melodies heard in a variable form.

When Halpern (1989) required nonmusicians to hum or sing the first note of familiar songs such as “Yankee Doodle” on two different occasions (separated by two days), the pitch varied minimally ($SD < 1.5$ semitones; see also Bergeson & Trehub, 2002). When participants were required instead to play the starting tone on a keyboard, there was more variability (implicating vocal limitations and/or motor memory in the initial experiment), but the SD was still less than three semitones between testing sessions. In other words, familiar songs like “Yankee Doodle” can be recognized in a wide range of keys, but a restricted pitch range (five to six semitones) is associated with canonical renditions.

Researchers have also exposed listeners to a set of novel melodies followed by a recognition test that includes the same (old) melodies as well as new melodies. By changing some of the old melodies in key, tempo, or timbre, long-term memory for surface features is implicated when recognition is enhanced for old melodies that remain identical from exposure to test. This paradigm, adopted here, capitalizes on the *encoding specificity principle* (Tulving & Thomson, 1973), which states that memory is enhanced when the context at test matches the context at exposure. In other words, memory for a melody should be improved if it is re-presented in the original key, tempo, or timbre, provided that listeners also remember these surface features.

In line with this view, Halpern and Müllensiefen (2008) found that memory for previously unfamiliar melodies was impacted negatively when the timbre changed from exposure to test, either from piano to organ (or vice versa) or from recorder to banjo (or vice versa). When these same researchers changed half of the old melodies in tempo by 15%–20% from exposure to test, memory for the changed melodies was again reduced. Peretz, Gaudreau, and Bonnel (1998, Exp. 3) found a similar detrimental effect of a surface change on recognition when melodies were changed from flute to piano (or vice versa) at test. When Lim and Goh (2012) changed melodies to a *similar* timbre at test (e.g., from violin to cello or vice versa), however, recognition was as good as when the melody was re-presented in the original timbre. This result implies either that memory for timbre is approximate or that a similar timbre provides a match between the test melody and the mental representation that is strong enough to confer the same recognition benefit as the original timbre. Other studies have provided converging evidence of memory for timbre (Wolpert, 1990) and tempo (Bergeson & Trehub, 2002), and also have shown that memory for melodies presented in some timbres (i.e., the human voice) is better than memory for melodies presented in other timbres (instruments; Weiss, Schellenberg, Trehub, & Dawber, 2015a; Weiss, Trehub, & Schellenberg, 2012; Weiss, Vanzella, Schellenberg, & Trehub, 2015b). Melody recognition is also impaired when the articulation format (i.e., legato or staccato) changes from exposure to test (Lim & Goh, 2013).

Studies of memory for the key of melodies have typically focused on short-term rather than long-term memory, using same–different tasks or similarity ratings for melodies presented one after the other (e.g., Bartlett & Dowling, 1980; Stalinski & Schellenberg, 2010). When listeners are asked to compare such standard and comparison melodies, increasing the pitch distance (i.e., the magnitude of the transposition in semitones) makes melodies sound dissimilar, as does increasing the key distance (i.e., fewer overlapping pitch classes; van Egmond & Povel, 1996; van Egmond, Povel, & Maris, 1996).

To date, only one study has tested long-term memory for the key of previously unfamiliar melodies (Schellenberg,

Stalinski, & Marks, 2014). To compare memory for key with memory for tempo, the researchers also included a tempo change designed to be equivalent in psychological magnitude to the key change. During the exposure phase, listeners heard each of 12 melodies twice. After a delay of approximately 10 min, they heard the same (old) melodies as well as 12 new melodies. Their task was to determine whether the melodies were old or new. Half of the old melodies were transposed by six semitones, changed in tempo by 64 beats per minute (bpm), or transposed *and* changed in tempo, but listeners were instructed specifically to ignore changes in key and/or tempo. Recognition of old melodies was excellent, but was even better when the surface features were unchanged. Although both key and tempo changes had detrimental effects on recognition, these effects were additive (no interaction) and similar in size. In short, listeners exhibited long-term memory for key and tempo after hearing a previously unfamiliar melody twice, and key and tempo appeared to be stored independently in listeners' mental representations.

Timbre may be a particularly salient surface feature because it provides information about the source of the melody, and because source cues tend to be encoded in mental representations of auditory objects (Johnson, Hashtroudi, & Lindsay, 1993). In one instance, nonmusicians judged two different melodies presented in the same timbre as being more similar than the same melody presented in two different timbres (Wolpert, 1990). In an analogous task with key instead of timbre changes, nonmusicians judged two different melodies presented in the same pitch range to be less similar than the same melody present in two different keys (Stalinski & Schellenberg, 2010). It is possible, then, that timbre may be retained significantly longer than key and/or tempo.

In the present investigation, we compared the time course of memory for melodies that remained identical from exposure to test to that for melodies shifted in key, tempo, or timbre. Listeners were asked whether they recognized melodies that they had heard 10 min, 1 day, or 1 week earlier, and told specifically to ignore changes in key, tempo, or timbre. In one previous study, researchers examined long-term memory for previously unfamiliar melodies and found that recognition deteriorated when the delay between exposure and test was increased from 1 day to 1 month, although melodies were remembered at above-chance levels in both cases (Peretz et al., 1998, Exp. 2). No evidence of reduced recognition emerged between delays of 5 min and 1 day.

On the basis of the available literature, our predictions were that mental representations of melodies would become more abstract over time, such that eventually they would be comprised primarily of relational information, and that memory for melodies would fade minimally over the course of a week, even in the presence of a change in key, tempo, or timbre. By contrast, memory for key and tempo was expected to fade

with increasing delay between exposure and test, such that re-presenting the melodies in the original key or tempo would no longer improve recognition. Memory for timbre, however, was expected to be longer lasting than memory for key and tempo. Musically trained listeners have exhibited enhanced memory for melodies in some studies but not others, although music training does not appear to influence memory for surface features (Dowling, Kwak, & Andrews, 1995; Schellenberg et al., 2014; Weiss et al., 2012, 2015b). Thus, we had no predictions regarding music training in the present study, except that it would not affect the recognition advantage for melodies that were identical at exposure and test.

Because key (pitch distance) and tempo (speed) are continuous variables, they were examined jointly and compared in Experiment 1, using manipulations of key and tempo that were designed to be similar in psychological magnitude (Schellenberg et al., 2014). Timbre comprises many different continuous dimensions (e.g., rise time, spectral center, spectral flux, and decay time; McAdams, 2013), however, such that instruments from different classes (e.g., piano and saxophone) are more appropriately considered as a categorical change. Hence, memory for melodies presented in different timbres was examined separately in Experiment 2.

Experiment 1

Method

Participants The sample comprised 192 undergraduates recruited without regard to music training. Because the apparatus, stimuli, and procedure (except for the delay) were identical to those used by Schellenberg et al. (2014, Exp. 1), the 64 listeners in the key- and tempo-change conditions from the earlier experiment—with a 10-min delay—were included in the sample. The new 128 listeners were tested with a 1-day or 1-week delay. On average, the participants had had 4.6 years of music lessons ($SD = 6.5$ years), but the distribution was skewed positively, as it tends to be in samples of undergraduates. In the statistical analyses, music training was treated as a binary variable, with 104 trained participants (≥ 2 years of lessons) and 88 untrained participants (< 2 years), as in previous studies with undergraduate samples (e.g., Dowling et al., 1995; Dowling, Tillmann, & Ayers, 2001; Schellenberg et al., 2014; Weiss et al., 2012).

Materials The stimuli were 24 melodies of approximately 30 s (12–16 measures, three or four phrases) taken from British and Irish folk-song collections, so that they were unfamiliar but tonal (Western-sounding), in major or minor mode, and in duple (4/4) or triple (3/4) meter. The average melody had tones with five different durations and nine

different pitches. Sixteen of the melodies were in one of five different major keys; the other eight melodies were in one of six different minor keys. The stimuli were entered note by note using Finale NotePad 2010 (MakeMusic Inc.). GarageBand 5.1 (Apple Inc.) was used to assign melodies to a piano timbre (Grand Piano), to change key or tempo, and to save the stimuli as digital (MP3) sound files. Customized software presented the stimuli over high-quality headphones and recorded the responses.

Procedure The procedure was identical to the recognition task used by Schellenberg et al. (2014), except that the new listeners had a delay of 1 day or 1 week between exposure and test, in contrast to the 10-min delay used in the previous study. Of the new participants, 64 were assigned to conditions with key changes, and 64 were assigned to conditions with tempo changes. Participants were then further subdivided into two groups of 32, based on the delay between exposure and test (1 day or 1 week).

Before the test session began, all participants heard different versions of “Happy Birthday” (high, low, fast, and slow) to demonstrate that key and tempo are irrelevant to a tune’s identity. All participants acknowledged quickly that they understood the point. In the exposure phase, participants heard 12 stimulus melodies: four in major mode and duple meter (4/4), four in major mode and triple meter (3/4), three in minor/duple, and one in minor/triple. They were required to make a happy/sad judgment to ensure that they had listened to each. In the key-change conditions, half of the melodies had a median pitch (adjusted for duration) of G4 (the G above middle C). The others had a median pitch of C#5 (six semitones higher). Melodies were equated for median pitch rather than key, so that the distinctiveness of any high or low notes would be minimized, and to ensure that the stimulus set comprised melodies in a variety of major and minor keys. (Equating on the basis of average pitch would have been influenced unduly by unusually high or low tones, and the average would not have corresponded to a pitch in the equal-tempered scale with A4 = 440 Hz.) The tempos were identical across melodies (110 bpm), which were presented in random order, followed by a second presentation in a different random order. The tempo-change conditions were identical, except that half of the melodies had a tempo of 110 bpm, half had a tempo of 174 bpm, and the median pitch was identical across melodies (G4). The tempo change represented a change of 58 % or 37 %, depending on whether the standard is considered to be 110 or 174 bpm, respectively.

During the test phase, which occurred after the delay, participants heard the 12 old melodies plus an additional 12 new melodies configured identically (re: mode, meter, key, and tempo). Their task was to rate on a 7-point scale whether they had heard the melody previously in the exposure

phase (1 = *Completely sure I didn’t hear that tune before*, 4 = *Not sure whether I heard that tune before*, 7 = *Completely sure I heard that tune before*). In the pitch-change conditions, half of the high or low old melodies were transposed down or up, respectively, by six semitones. In the tempo-change conditions, half of the fast or slow melodies were slowed down or sped up, respectively, by 64 bpm. In all conditions, participants were instructed specifically to ignore key and tempo changes and were reminded that key and tempo are irrelevant to a melody’s identity. Melodies were counterbalanced across listeners, so that all were presented equally often as old or new, high or low (or fast or slow), and changed or unchanged from exposure to test. The orders of presentation were randomized separately for the exposure and test phases, and separately for each participant.

Results and discussion

Recognition ratings were converted to scores measuring area under the receiver operating characteristic (ROC) curve, which we refer to as *area under the curve* (AUC), as in previous studies that had examined memory for music (e.g., Dowling et al., 1995; Dowling et al., 2001). (Analyses of d' scores and the raw data are provided in the [supplementary materials](#). The three approaches produced very similar results.) AUC scores were calculated from the ratings for old melodies (hits) and new melodies (false alarms), to provide unbiased estimates of proportions correct (Swets, 1973). AUC scores are considered to be less biased than d' scores (Dowling et al., 1995; Verde, Macmillan, & Rotello, 2006), and they do not require transformation of the original ratings into a dichotomy, such that they retain more information from the original rating scale. They range in principle from 0 to 1.0, with chance being equal to .5 (i.e., responding identically to new and old melodies) and perfect responding equal to 1.0 (i.e., all ratings for old melodies higher than all ratings for new melodies).

Three AUC scores were formed for each participant: one for all old melodies, another for old-same melodies, and a third for old-changed melodies. The overall score was calculated from the 24 data points for each participant (12 hit rates and 12 false alarm rates), whereas the scores for old-same and old-changes melodies were calculated from 18 data points (6 hit rates and 12 false alarm rates). All scores exceeded chance levels at all delays, $ps < .001$.

Initial analyses focused on the overall AUC scores. Descriptive statistics are provided in Table 1 separately for each of the six groups of participants (3 delays \times 2 domains). Between-condition differences were analyzed with an analysis of variance (ANOVA) that had three between-subjects factors: Domain (pitch or tempo) and Delay (10 min, 1 day, and 1 week), as well as Music Training (trained or untrained). We found no main effect of music training and no interactions

Table 1 Descriptive statistics for recognition, as measured by area under the curve in Experiment 1

Domain	Delay	Overall		Old-Same		Old-Changed	
		<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Key	10 min	.816	.086	.859	.117	.773	.112
	1 day	.863	.098	.890	.103	.835	.129
	1 week	.854	.104	.860	.143	.849	.102
Tempo	10 min	.815	.109	.855	.111	.765	.143
	1 day	.844	.099	.888	.086	.800	.140
	1 week	.881	.106	.891	.116	.870	.119

involving music training, $p > .3$. There was a small but significant main effect of delay (Fig. 1), $F(2, 180) = 5.03, p = .007, \eta_p^2 = .053$, but no main effect of domain, $F < 1$, and no two-way interaction between domain and delay, $p > .3$. Follow-up tests of the delay effect (Tukey’s HSD) revealed surprising results: Melody recognition improved as the delay increased. Specifically, AUC scores were higher after a 1-week delay than after a 10-min delay, $p = .010$, and marginally higher after a 1-day delay than after a 10-min delay, $p = .087$. Recognition was similar after a 1-day or 1-week delay, $p > .6$.

We then examined whether recognition was influenced by changes in key or tempo, and whether such influences varied as a function of domain, the delay between initial exposure and test, and music training. Descriptive statistics are reported in Table 1 as a function of domain, delay, and surface feature. A mixed-design ANOVA with one repeated measure (Surface Feature: same or changed) and three between-subjects variables (Delay, Domain, and Music Training) showed no main effects of domain or music training and no interactions involving domain or music training, $ps > .2$. A robust main effect of surface feature, $F(1, 180) = 37.03, p < .001, \eta_p^2 = .171$, confirmed that melodies were recognized better if the key and tempo remained identical from exposure to test. As in the first

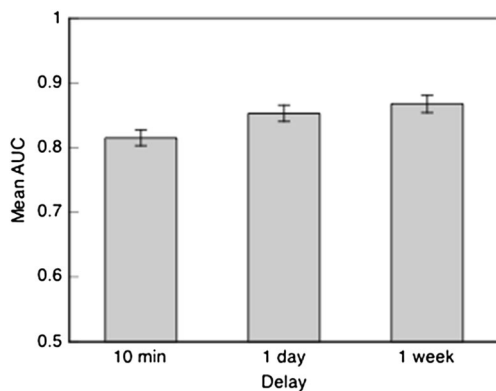


Fig. 1 Melody recognition in Experiment 1 (collapsed across key and tempo conditions) as a function of the delay between exposure and test. Recognition improved as the delay increased. The data in the 10-min delay conditions are from Schellenberg et al. (2014), and error bars are standard errors of the means

analysis, there was a main effect of delay, $F(2, 180) = 5.44, p = .005, \eta_p^2 = .057$, but a two-way interaction between surface feature and delay, $F(2, 180) = 5.76, p = .004, \eta_p^2 = .060$, qualified both main effects (Fig. 2). Melodies in the original key and tempo were recognized better than changed melodies after a 10-min delay, $F(1, 60) = 24.13, p < .001, \eta_p^2 = .287$, and a 1-day delay, $F(1, 60) = 21.29, p < .001, \eta_p^2 = .262$, but not after a 1-week delay, $F < 1$.

In additional follow-up analyses, we examined the effect of the delay separately for old-same and old-changed melodies. For old-same melodies, we found no effect of delay, $p > .3$. For old-change melodies, the effect of the delay, $F(2, 180) = 9.28, p < .001, \eta_p^2 = .093$, stemmed from better recognition after a 1-week delay than after 10 min, $p < .001$, and marginally better recognition after a 1-day delay than after 10 min, $p = .076$. Old-changed melodies were recognized similarly after a 1-day or 1-week delay, $p > .1$. In short, the increase in overall recognition as the delay increased was a consequence of better recognition of old-changed melodies.

To summarize, the analyses showed that (1) melodies were recognized well after all delays; (2) there was no sign of a reduction in recognition as the delay increased; (3) recognition of old-changed melodies improved over time; (4) re-presenting a melody in the original key or tempo improved recognition after a short or medium delay (10 min or 1 day), but not after a long delay (1 week); (4) the disruptive effects of the key and tempo changes were similar; and (5) response patterns were independent of music training.

Experiment 2

Experiment 2 was identical to Experiment 1, except that the surface feature change involved a change in timbre. As we

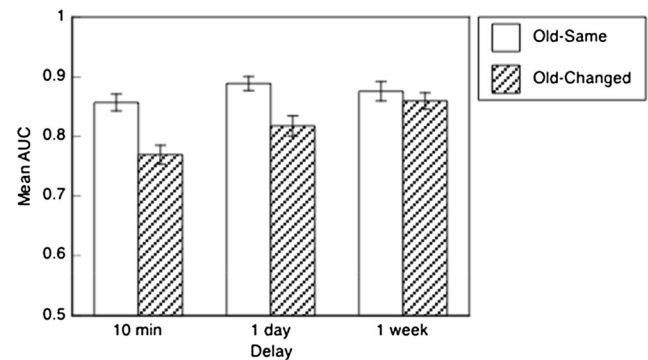


Fig. 2 Melody recognition in Experiment 1 (collapsed across key and tempo conditions) as a function of the delay between exposure and test, and whether the melodies underwent a change in key or tempo. The key or tempo change negatively affected recognition after a 10-min or 1-day delay, but not after a 1-week delay. The data in the 10-min conditions are from Schellenberg et al. (2014), and error bars are standard errors of the means

noted, there is no obvious way to manipulate timbre on a single dimension that would make it similar in psychological magnitude to the manipulations of key and tempo in Experiment 1, or to an actual change from one musical instrument to another. Instead, the timbre change in the present experiment—from piano to saxophone, or vice versa—was designed to be ecologically valid yet obvious, much like the key and tempo manipulations in Experiment 1.

Method

Participants The participants were 96 undergraduates, recruited without regard to music training, as in Experiment 1. The average duration of music training was similar to that in Experiment 1 ($M = 5.3$ years, $SD = 7.4$) and was positively skewed. In the statistical analyses, music training was again considered a binary variable, with 53 trained participants (≥ 2 years of lessons) and 43 untrained participants (< 2 years).

Materials These were identical to the materials of Experiment 1, except that the change from exposure to test involved a change in timbre—from piano to saxophone (alto sax) or vice versa—rather than a change in key or tempo. Thus, the stimuli for the present experiment included 24 different melodies in a piano timbre and the same 24 melodies in a saxophone timbre. Melodies were presented at the lower pitch and slower tempo from Experiment 1.

Procedure This was again as in Experiment 1, except that the initial demonstrations involved “Happy Birthday” presented in different timbres. All participants readily understood that a change in timbre is irrelevant to a melody’s identity. Before the recognition test, all participants were again reminded of this fact.

Results and discussion

Recognition scores were derived as in Experiment 1, with AUC scores calculated for overall recognition, recognition of old-same melodies, and recognition of old-changed melodies. Performance was above chance levels at each of the three delays for each of the three scores, $ps < .001$. An ANOVA with two between-subjects factors (Delay and Music Training) revealed that overall melody recognition was better for musically trained than for untrained participants, $F(1, 90) = 5.50$, $p = .021$, $\eta_p^2 = .058$. The delay did not affect overall recognition, $p > .1$, and no interaction was apparent between music training and delay, $F < 1$. Descriptive statistics are illustrated in Fig. 3.

A mixed-design ANOVA with one repeated measure (Timbre: same or changed) and two between-subjects factors (Delay, Music Training) revealed a main effect of timbre, $F(1,$

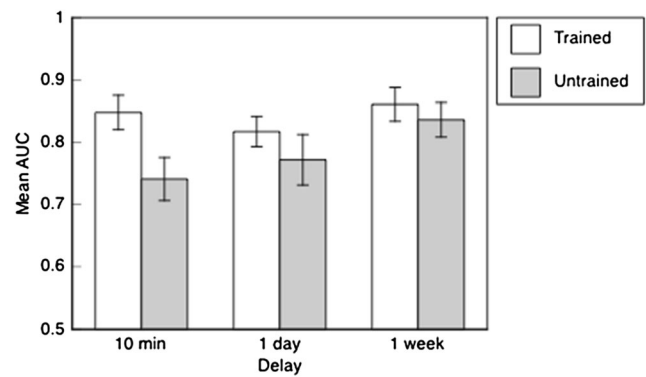


Fig. 3 Melody recognition in Experiment 2 as a function of music training and the delay between exposure and test. Musically trained individuals performed better than their untrained counterparts, but there was no main effect of delay and no interaction between music training and delay. Error bars are standard errors of the means

$90) = 19.70$, $p < .001$, $\eta_p^2 = .180$ (Fig. 4): Recognition was better for melodies re-presented in the original timbre than for melodies that changed timbre from exposure to test. The effect of the delay was not significant, $p > .1$, and no interaction emerged between timbre and delay, $p > .2$. In other words, maintaining the same timbre at exposure and test facilitated recognition to a similar degree after all delays. The main effect of music training was again evident, $F(1, 90) = 5.37$, $p = .023$, $\eta_p^2 = .056$, but music training did not interact with timbre or delay, $F_s < 1$, which confirmed that the recognition advantage for same-timbre melodies was evident regardless of music training.

Thus, the analyses revealed three main results: (1) Memory for melodies was unaffected by the delay between exposure and test, (2) re-presenting melodies in the original timbre led to similarly enhanced recognition after each of the three delays, and (3) musically trained participants had better recognition than musically untrained participants after all delays, even though both groups showed similar decrements in recognition as a consequence of the timbre change.

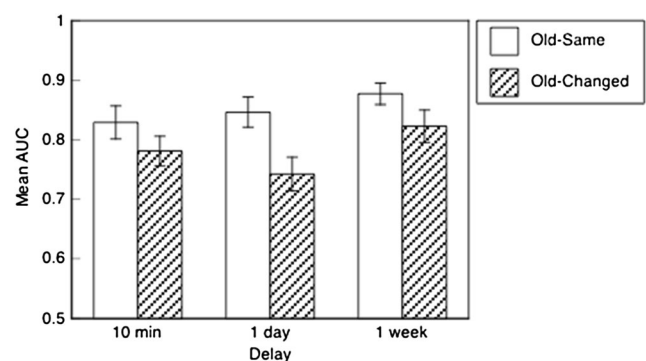


Fig. 4 Melody recognition in Experiment 2 as a function of the delay between exposure and test, and whether the melodies underwent a change in timbre. The timbre change negatively affected recognition after each delay. Error bars are standard errors of the means

General discussion

In two experiments, memory for the pitch and temporal relations that defined specific melodies was robust across delays, with no sign of forgetting even after the longest (1-week) delay. In Experiment 1, re-presenting the melodies in the original key or tempo improved recognition after a 10-min and 1-day delay, but not after 1 week. In Experiment 2, melodies heard in the same timbre at exposure and test were recognized better than melodies that changed timbre from exposure to test, and the effect was similar after all delays.

The present results are notable for at least five reasons. First, they reveal that memory for the relational information that defines melodies appears to be fully consolidated after two exposures, such that no forgetting occurs over a 1-week interval. Second, they document that the surface features of such melodies are stored in listeners' mental representations much longer than 1 min, contrary to long-held beliefs (Krumhansl, 2000; Raffman, 1993), and well beyond 10 min (Schellenberg et al., 2014). Third, evidence of long-term memory for key, tempo, and timbre is apparent in unselected listeners rather than being limited to highly trained individuals or those with absolute pitch. Fourth, the findings highlight different trajectories of forgetting for abstract information that defines a melody's identity and for surface information associated with specific renditions. Fifth, some surface features serve as recognition cues for longer durations than other surface features.

Although one might question whether the key and tempo manipulations were too subtle to allow the original key and tempo to be used as cues for recognition after a 1-week delay, the changes (six semitones, 64 bpm) were large musically and psychophysically. To illustrate, upward and salient key changes that occur near the end of a song to create surprise are usually only one or two semitones. In Whitney Houston's "I Wanna Dance With Somebody (Who Loves Me)," the transposition is one semitone. In the same singer's "I Will Always Love You" (from the film *The Bodyguard*), a more dramatic transposition, which occurs after a moment of silence, is two semitones. In both instances, the key changes are salient, perhaps even jarring to some listeners, yet one-third of the magnitude of the key manipulations in the present investigation. In the case of tempo, a change of 64 bpm is about 15 times greater than the JND for tempo discrimination when the stimuli involve actual music (Levitin & Cook, 1996).

One notable and provocative result from Experiment 1 was that recognition of old-changed melodies actually improved as the delay between exposure and test increased from 10 min to 1 week. This finding suggests that at the shortest delay, listeners' mental representations contained information about key and tempo, such that the mismatch between the representation and the test stimulus was highlighted. Over time, as key and tempo information faded from the representation, only

relational information was retained, such that the test stimulus matched the representation. According to this view, memory for the surface features of a melody—key and tempo—interfered with memory for abstract features at the shortest delay.

With an even larger sample or a different method, a recognition advantage after a 1-week delay for melodies presented in the original key and tempo could be significant, although small in magnitude. To illustrate, if we assume that the real difference between the old-same and old-changed conditions is actually twice as large as the one we observed (i.e., a difference in mean AUC of .03256 instead of .01628), a sample of over 400 would be required to have an 80 % chance of rejecting the null hypothesis (Howell, 2013). In a classic study of memory (Tulving, Schacter, & Stark, 1982), participants did not recognize words seen 1 week earlier, yet performance on a priming task (i.e., word-fragment completion) was enhanced by previous exposure, which raises the possibility that a different task could lead to different findings. Response patterns could also differ if popular recordings were used as the stimuli. Because digital recordings are heard repeatedly at exactly the same key and tempo, changes in key or tempo would likely be noticeable for longer durations than they were for our previously unfamiliar stimulus melodies.

In Experiment 2, re-presenting the melodies in the original timbre improved recognition at all delays. In absolute terms, after a 1-week delay, the difference in recognition confidence between old-same and old-changed melodies was actually larger (.054) than after a 10-min delay (.048; Fig. 4), which implies that statistical power was a moot point. Why do timbre manipulations have such an enduring effect on melody recognition? As we noted in the introduction, timbre is a source cue, more or less analogous to individual voices in the case of speech. Speakers can speak in a high or a low tone of voice, or rapidly or slowly, yet their identity remains constant. Timbre identifies *who* or *what*, whereas pitch height (key) and tempo are more likely to depend on situational factors. Key and tempo changes in music (with timbre held constant) are comparable to the same speaker using a high or low register, or a faster or slower rate. More generally, discriminating sources has more relevance for survival than does discriminating different aspects of the same source.

In Experiment 1, recognition of melodies that underwent a change in key or tempo actually improved as the delay increased. In Experiment 2, recognition was stable across delays, presumably because the timbre change had similar detrimental effects at all delays. How can we account for listeners' excellent memory of the melodies, such that listeners showed no signs of forgetting over time? Conformity to Western tonal and metrical structures undoubtedly played a role. For example, the notes of each melody came from a single key, with a clearly defined reference tone (the note called *doh*), and all of the melodies were in a familiar (duple or triple) meter. In another study that used unfamiliar but

Western-sounding melodies, memory faded over time but was still above chance levels after 1 month (Peretz et al., 1998). Although listeners with limited music training lack explicit knowledge of mode or meter, their implicit knowledge of Western tonal and metrical structures would have facilitated perception (e.g., Hannon & Trehub, 2005; Lynch, Eilers, Oller, & Urbano, 1990; Trehub & Hannon, 2009) and memory (Demorest, Morrison, Beken, & Jungbluth, 2008; Demorest et al., 2010; Wong, Roy, & Margulis, 2009) in the present study. Had the melodies been atonal, nonmetrical, and/or drawn from a foreign musical system (e.g., Balinese, Balkan, Indian), recognition would have been much more difficult, and it is unknown whether changes in surface features would have a detrimental effect similar to the ones that we observed.

Musically trained participants exhibited better memory for melodies in Experiment 2, but not in Experiment 1. Similar inconsistencies were evident in previous studies (e.g., Dowling et al., 1995; Schellenberg et al., 2014; Weiss et al., 2012, 2015b). Indeed, when the present data were analyzed in terms of d' scores (see the [supplementary materials](#)), musically trained participants showed no memory advantage in either experiment. When the raw data were analyzed ([supplementary materials](#)), on the other hand, musically trained individuals exhibited a recognition advantage in both experiments. Advantages on melodic-memory tasks for musically trained individuals may be more reliable when comparing musicians with many years of lessons to individuals with no formal training (Weiss et al., 2015b). In any event, the detrimental effects of changes in surface features on recognition memory were independent of music training in both experiments and in all analyses.

The present findings have parallels with language research. Abstractionist linguistic theories (e.g., Halle, 1985) posit that spoken words are *normalized*—stripped of surface features (e.g., speaker's voice)—before being compared with abstract mental representations in the lexicon. We now know, however, that listeners process words and voices in tandem, such that recognition of previously heard words is best when the words are produced by the same speaker at exposure and test (e.g., Goh, 2005; for a review, see Nygaard, 2005). As with music, then, listeners remember the abstract *and* surface features of spoken words.¹

¹ The distinction between surface and abstract features in music and language that we make here is similar to that used in other studies of music (Lim & Goh, 2012, 2013; Plantinga & Trainor, 2005; Poulin-Charronnat et al., 2004; Schellenberg et al., 1999; Schellenberg & Trehub, 2003, 2008; Trainor et al., 2004; Trehub et al., 2008; Volkova et al., 2006) and language (e.g., Goh, 2005; Goldinger, 1996; Nygaard, 2005). It differs, however, from the surface/abstract distinction made by Dowling, Tillmann, and colleagues (e.g., Tillmann & Dowling, 2007; Tillmann et al., 2013), who used *surface* to describe the exact intervals in music and the words in language, and *abstract* to describe the contour and rhythm in music and the meaning in language.

Goldinger (1996, Exp. 2) examined the duration of voice-specific long-term memory by testing the recognition of monosyllabic words soon after exposure, 1 day later, or 1 week later. Recognition was good at all delays, with a hit rate of approximately 65 % correct or better (yes–no task, chance = 50 %), even when the speaker changed from exposure to test. At the shortest delay (two-speaker condition, analogous to two keys or two tempi), word recognition was enhanced by 7.6 % when the speaker was the same at exposure and test. After 1 day, the same-speaker advantage was reduced to 5.7 %, and the advantage disappeared after 1 week (1.3 %). On the one hand, the similarity of Goldinger's response patterns with those reported here for key and tempo raises the possibility that the time courses of memory for some surface features of language and music are comparable. On the other hand, Goldinger's null effect after a 1-week delay contrasts with our evidence of memory for timbre after the same delay, particularly because different timbres in music are directly analogous to different voices in speech.

Goldinger (1996, Exp. 2) considered his test of voice-specific word recognition—similar to the present tests of key-, tempo-, and timbre-specific melody recognition—to be one of *explicit* memory, even though participants were *not* required to identify whether the speaker had changed, and the instructions told them to ignore this manipulation, as in the present experiments and other previous research (Goh, 2005; Lim & Goh, 2012, 2013; Schellenberg et al., 2014). For a separate group of listeners who were required to identify words embedded in white noise, identification performance was 6 % better for words heard 1 week earlier in the same voice, which unambiguously implicated voice-specific implicit memory, because there were no judgments of recognition. In the present experiments and in Goldinger's word recognition task, however, the implicit–explicit distinction is not completely clear. Because participants were not required to identify whether the stimuli at test were identical to those at exposure, the measures of memory for voice, key, tempo, or timbre appear to be implicit. Because listeners were required to make recognition judgments rather than judgments in a domain distinct from recognition, such as liking, the measures appear to be explicit. The implicit–explicit distinction could be explored in future research by requiring participants to judge whether each melody at test is identical in all respects (re: key, tempo, and timbre) to a melody presented at exposure (an explicit task), or by requiring participants to make liking judgments in the test phase (an implicit task).

In audition as well as vision, a central problem is explaining generalization, or how different sensory experiences (e.g., faces viewed from different angles, words said by different speakers, or melodies presented in different keys/tempi/timbres) give rise to perception of the same object. In the case of melodies, memory for surface features represents an instance of *feature binding* (Treisman & Gelade,

1980), such that mental representations contain information about key, tempo, and timbre, as well as pitch intervals and rhythm. Unlike vision, however, in which a common location provides a cue that features should be bound into a single object, auditory surface and abstract features are bound together because they occur in synchrony as they unfold over time. In the case of melodies, perhaps the most interesting phenomenon is one of asymmetrical *unbinding* over time between abstract and some surface features, such as key and tempo. The abstract information is retained, whereas memory for surface cues fades unless the material is heard repeatedly with the identical cluster of abstract and surface cues (e.g., Levitin & Cook, 1996; Schellenberg & Trehub, 2003). Timbre, by contrast, appears to be bound with the abstract features for a longer duration, perhaps indefinitely.

Previous research has examined melodic binding solely in terms of abstract features, with the time between initial exposure and test being comparable to our shortest delay. Exact pitch intervals in music (Dowling, 1991; Dowling & Bartlett, 1981; Dowling et al., 1995) are increasingly bound to the underlying rhythm and contour structure over time, such that there is improvement in memory (or no forgetting) over the course of 15–120 s. The same phenomenon is observed with homophonic music (i.e., melody plus harmony; Dowling & Bartlett, 1981; Dowling & Tillmann, 2014; Dowling et al., 2001; Tillmann et al., 2013). Memory for exact words in poetry shows similar improvements, presumably because the words are increasingly bound to the underlying rhythm over time, which explains why the effect is not evident for prose (Tillmann & Dowling, 2007).

The present findings are consistent with previous results showing that infants, children, and adults remember pitch relations (i.e., contour, intervals) and absolute pitch level or key (for a review, see Stalinski & Schellenberg, 2012), but only individuals with absolute pitch remember *labels* for specific tones, which allows them to produce or label a musical tone in isolation. Nevertheless, there is an ontogenetic shift in focus from pitch level or key to pitch relations. For example, a melody and its transposition are considered virtually identical by adults, but not by children 12 years of age or younger (Stalinski & Schellenberg, 2010). For children under 10 years of age, a comparison melody created by reordering the tones of a standard melody seems similar to the standard, even though all pitch relations have changed, yet simply transposing the standard makes it markedly different. There is a parallel phylogenetic shift, with songbirds being more likely than humans to remember key (Weisman, Williams, Cohen, Njegovan, & Sturdy, 2006). The present findings document a similar shift based on delay after exposure that extends beyond key to tempo, but not to timbre. In any event, over development, evolution, and time since exposure, melodic memory becomes increasingly abstract, which helps to explain why music, like language, is a uniquely human experience.

Author note This study was funded by the Natural Sciences and Engineering Research Council of Canada, and assisted by Katie Corrigan, Safia Khalil, and Rusan Lateef. Bill Thompson and Sandra Trehub provided helpful comments on an earlier draft.

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
- Demorest, S. M., Morrison, S. J., Beken, M. N., & Jungbluth, D. (2008). Lost in translation: An enculturation effect in music memory performance. *Music Perception*, 25, 213–223. doi:10.1525/mp.2008.25.3.213
- Demorest, S. M., Morrison, S. J., Stambough, L. A., Beken, M., Richards, T. L., & Johnson, C. (2010). An fMRI investigation of the cultural specificity of music memory. *Social Cognitive and Affective Neuroscience*, 5, 282–291. doi:10.1093/scan/nsp048
- Deutsch, D. (2013). Absolute pitch. In D. Deutsch (Ed.), *The psychology of music* (3rd ed., pp. 141–182). San Diego, CA: Elsevier Academic Press. doi:10.1016/B978-0-12-381460-9.00005-5
- Dowling, W. J. (1991). Tonal strength and melody recognition after long and short delays. *Perception & Psychophysics*, 50, 305–313. doi:10.3758/BF03212222
- Dowling, W. J., & Bartlett, J. C. (1981). The importance of interval information in long-term memory for melodies. *Psychomusicology*, 1, 30–49.
- 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
- 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
- Goh, W. D. (2005). Talker variability and recognition memory: Instance-specific and voice-specific effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 31, 40–53. doi:10.1037/0278-7393.31.1.40
- Goldinger, S. D. (1996). Words and voices: Episodic traces in spoken word identification and recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22, 1166–1183. doi:10.1037/0278-7393.22.5.1166
- Halle, M. (1985). Speculations about the representation of words in memory. In V. Fromkin (Ed.), *Phonetic linguistics* (pp. 101–114). New York, NY: Academic Press.
- Halpern, A. R. (1989). Memory for the absolute pitch of familiar songs. *Memory & Cognition*, 17, 572–581. doi:10.3758/BF03197080
- 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

- Hannon, E. E., & Trehub, S. E. (2005). Metrical categories in infancy and adulthood. *Psychological Science*, *16*, 48–55. doi:10.1111/j.0956-7976.2005.00779.x
- Howell, D. C. (2013). *Statistical methods for psychology* (8th ed.). Belmont, CA: Wadsworth.
- Johnson, M. K., Hashtroudi, S., & Lindsay, D. S. (1993). Source monitoring. *Psychological Bulletin*, *114*, 3–28. doi:10.1037/0033-2909.114.1.3
- Krumhansl, C. L. (2000). Rhythm and pitch in music cognition. *Psychological Bulletin*, *126*, 159–179. doi:10.1037/0033-2909.126.1.159
- 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
- Lynch, M. P., Eilers, R. E., Oller, D. K., & Urbano, R. C. (1990). Innateness, experience, and music perception. *Psychological Science*, *1*, 272–276. doi:10.1111/j.1467-9280.1990.tb00213.x
- McAdams, S. (2013). Musical timbre perception. In D. Deutsch (Ed.), *The psychology of music* (3rd ed., pp. 35–67). San Diego, CA: Elsevier. doi:10.1016/B978-0-12-381460-9.00002-X
- Nygaard, L. C. (2005). Perceptual integration of linguistic and nonlinguistic properties of speech. In D. B. Pisoni & R. E. Remez (Eds.), *Handbook of speech perception* (pp. 390–413). Malden, MA: Blackwell.
- 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
- Poulin-Charronnat, B., Bigand, E., Lalitte, P., Madurell, F., Vieillard, S., & McAdams, S. (2004). Effects of a change in instrumentation on the recognition of musical materials. *Music Perception*, *22*, 239–263. doi:10.1525/mp.2004.22.2.239
- Raffman, D. (1993). *Language, music, and mind*. Cambridge, MA: MIT Press.
- Randel, D. (Ed.). (1986). *The new Harvard dictionary of music*. Cambridge, MA: Belknap Harvard.
- 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., 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
- Smith, N. A., & Schmuckler, M. A. (2008). Dial A440 for absolute pitch: Absolute pitch memory by non-absolute pitch processors. *Journal of the Acoustical Society of America*, *123*, EL77–EL84. doi:10.1121/1.2896106
- 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
- Stalinski, S. M., & Schellenberg, E. G. (2012). Music cognition: A developmental perspective. *Topics in Cognitive Science*, *4*, 485–497. doi:10.1111/j.1756-8765.2012.01217.x
- Swets, J. A. (1973). The relative operating characteristic in psychology. *Science*, *182*, 990–1000.
- Tillmann, B., & Dowling, W. J. (2007). Memory decreases for prose, but not for poetry. *Memory & Cognition*, *35*, 628–639. doi:10.3758/BF03193301
- Tillmann, B., Dowling, W. J., Lalitte, P., Molin, P., Schulze, K., Poulin-Charronnat, B., & Bigand, E. (2013). Influence of expressive versus mechanical musical performance on short-term memory for musical excerpts. *Music Perception*, *30*, 419–425. doi:10.1525/mp.2013.30.4.419
- 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., & Hannon, E. E. (2009). Conventional rhythms enhance infants' and adults' perception of musical patterns. *Cortex*, *45*, 110–118. doi:10.1016/j.cortex.2008.05.012
- 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
- Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, *12*, 97–136. doi:10.1016/0010-0285(80)90005-5
- Tulving, E., Schacter, D. L., & Stark, H. A. (1982). Priming effects in word-fragment completion are independent of recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *8*, 336–342. doi:10.1037/0278-7393.8.4.336
- Tulving, E., & Thomson, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review*, *80*, 352–373. doi:10.1037/h0020071
- van Egmond, R., & Povel, D.-J. (1996). Perceived similarity of exact and inexact transpositions. *Acta Psychologica*, *92*, 283–295. doi:10.1016/0001-6918(95)00043-7
- van Egmond, R., Povel, D.-J., & Maris, E. (1996). The influence of height and key on the perceptual similarity of transposed melodies. *Perception & Psychophysics*, *58*, 1252–1259. doi:10.3758/BF03207557
- Verde, M. F., Macmillan, N. A., & Rotello, C. M. (2006). Measures of sensitivity based on a single hit rate and false alarm rate: The accuracy, precision, and robustness of d' , A_z , and A' . *Perception & Psychophysics*, *68*, 643–654. doi:10.3758/BF03208765
- Volkova, A., Trehub, S. E., & Schellenberg, E. G. (2006). Infants' memory for musical performances. *Developmental Science*, *9*, 583–589. doi:10.1111/j.1467-7687.2006.00536.x
- Weisman, R. G., Williams, M. T., Cohen, J. S., Njegovan, M. G., & Sturdy, C. B. (2006). The comparative psychology of absolute pitch. In E. A. Wasserman & T. R. Zentall (Eds.), *Comparative cognition: Experimental explorations of animal intelligence* (pp. 71–86). New York, NY: Oxford University Press.
- Weiss, M. W., Schellenberg, E. G., Trehub, S. E., & Dawber, E. J. (2015a). Enhanced processing of vocal melodies in childhood. *Developmental Psychology*, *51*, 370–377. doi:10.1037/a0038784
- Weiss, M. W., Vanzella, P., Schellenberg, E. G., & Trehub, S. E. (2015b). Pianists exhibit enhanced memory for vocal melodies but not piano melodies. *Quarterly Journal of Experimental Psychology*. doi:10.1080/17470218.2015.1020818
- Weiss, M. W., Trehub, S. E., & Schellenberg, E. G. (2012). Something in the way she sings: Enhanced memory for vocal melodies. *Psychological Science*, *23*, 1074–1078. doi:10.1177/0956797612442552

- Wolpert, R. A. (1990). Recognition of melody, harmonic accompaniment, and instrumentation: Musicians vs. nonmusicians. *Music Perception*, 8, 95–106. doi:10.2307/40285487
- Wong, P. C. M., Roy, A. K., & Margulis, E. H. (2009). Bimusicalism: The implicit dual enculturation of cognitive and affective systems. *Music Perception*, 27, 81–88. doi:10.1525/mp.2009.27.2.81