



UNIVERSITY OF CALIFORNIA PRESS
JOURNALS + DIGITAL PUBLISHING

Memory Improvement While Hearing Music: Effects of Structural Continuity on Feature Binding

Author(s): W. Jay Dowling and Barbara Tillmann

Source: *Music Perception: An Interdisciplinary Journal*, Vol. 32, No. 1 (September 2014), pp. 11-32

Published by: [University of California Press](#)

Stable URL: <http://www.jstor.org/stable/10.1525/mp.2014.32.1.11>

Accessed: 02/12/2015 13:16

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at <http://www.jstor.org/page/info/about/policies/terms.jsp>

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.



University of California Press is collaborating with JSTOR to digitize, preserve and extend access to *Music Perception: An Interdisciplinary Journal*.

<http://www.jstor.org>

MEMORY IMPROVEMENT WHILE HEARING MUSIC: EFFECTS OF STRUCTURAL CONTINUITY ON FEATURE BINDING

W. JAY DOWLING
University of Texas at Dallas

BARBARA TILLMANN
Lyon Neuroscience Research Center, Lyon, France

PREVIOUS RESEARCH (*MUSIC PERCEPTION*, 2002, Issue 2) demonstrated improvement in recognition memory across delays increasing from 5 to 15 s while listening to novel music, attributable to a decline in false alarms to similar lures. We hypothesize that this improvement results from delayed binding of features. At short delays, targets and similar lures are easily confused because they share individual features such as melodic contour and musical key. Binding those features into a coherent memory representation—such as encoding the pitch level at which the contour is attached to the scale—reduces that confusion and hence false alarms to similar lures. Here we report eight experiments in which we explore the conditions under which this continued encoding occurs, and test specific hypotheses concerning the particular features involved. These phenomena involve the binding of complex features of nonverbal material, and are explained in terms of theoretical descriptions of the features and the representations resulting from binding. We envisage future studies investigating this binding phenomenon with neurophysiological methods in the study of cognition in aging.

Received: August 3, 2012, accepted November 9, 2013.

Key words: memory, music, binding, features, contour

RECOGNITION MEMORY FOR MUSIC DIFFERS from verbal memory. For example, Sachs (1967) tested memory for sentences embedded in prose stories and found that memory for details, such as exact word order, declined with increasing delay. Sachs found that participants' ability to discriminate between repetitions of target sentences and similar lures (which had the same meaning as targets but different word order) declined over time. Tillmann and Dowling (2007, Experiment 3) replicated Sachs's result using

a format similar to that of the present experiments. Listeners heard prose stories in which target sentences that occurred early in the passage were tested after delays of 4.8 and 28.8 s during which they heard the continuation of the story. The results are shown in Table 1 and Figure 1. Discrimination between target test items and similar, same-meaning lures declined from 82% correct to 63% over the delay, as measured by area under the ROC (Receiver Operating Characteristic—a measure of unbiased proportion correct where chance equals .50, derived from plotting hit vs. false-alarm rates obtained used confidence-level response scales—Swets, 1973).¹ This result is reflected in the hit and false-alarm rates: hits (correct responses to target test items) decreased from .81 to .68 across the delay, while false alarms to similar lures increased from .26 to .38.² In contrast, Dowling, Tillmann, and Ayers (2002, Experiment 1; see Table 1 and Figure 2) tested memory for musical phrases embedded in pieces and found that memory for musical details improved with increasing delay, seen in the improvement in area scores from .61 to .78. At a short delay of 5 s, participants had difficulty rejecting similar lures (which shared the melodic and rhythmic contour of targets, but with the melody at a different pitch level), but were able to reject them after a longer delay of 15 s (over which time the false-alarm rate declined from .56 to .22). Thus, there was an opposite effect of delay on discrimination of targets from similar lures in memory for prose stories vs. memory for pieces of music.

This result was surprising not only because it showed a clear divergence between results with verbal materials and with music, but also because of its implications for

¹In our previous studies (Dowling et al., 2002; Tillmann & Dowling, 2007; Tillmann et al., 2013) and in the literature this is often referred to as MOC (Memory Operating Characteristic).

²The principle difference between Sachs's (1967) method and ours is that whereas our short delay included brief filler material (as in the music studies), hers involved an immediate test. This led to better performance at her short delay than ours. Our long delays were roughly the same as hers, and the hit and false-alarm rates are quite similar. In Sachs's Figure 1 the hit and false-alarm rates for test targets, similar lures (averaged over her two types), and different lures, for the short and long delay, were (respectively): .92-.64, .15-.39, and .12-.28. Note the similarity to the long delay results from Tillmann and Dowling (2007) in Table 1.

TABLE 1. Area Under the ROC for Target Test Item/Similar Lure (T/S) and Target Test Item/Different Lure (T/D), and Comparisons and Hits and False Alarms for Target Test Items (T), Similar Lures (S), and Different Lures (D) at Short and Long Time Delays in Experiments 1-8.

	Comparison				Item					
	T/S		T/D		T		S		D	
	Short	Long	Short	Long	Short	Long	Short	Long	Short	Long
Dowling et al. Exp 1	.61	.78	.78	.81	.73	.70	.56	.22	.26	.16
Dowling et al. Exp 3 (silent interval)	.87	.86	.97	.95	.87	.77	.35	.14	.07	.02
Tillmann & Dowling (Exp. 3; prose stories)	.82	.63	.78	.78	.81	.68	.26	.38	.29	.24
1: music (replication), $N = 44$.69	.76	.80	.80	.66	.42	.42	.24	.22	.17
2: scrambled beats within measures ($N = 26$)	.61	.71	.73	.75	.69	.60	.49	.32	.35	.24
3: filler transposed 1 semitone ($N = 37$)	.63	.75	.77	.81	.67	.67	.51	.27	.24	.18
4: gavotte filler – different key & meter ($N = 20$)	.66	.64	.81	.77	.79	.59	.56	.44	.31	.22
5: minuet filler – different key, same meter ($N = 35$)										
same minuet control	.48	.75			.58	.79	.63	.41		
different minuet	.65	.69			.61	.56	.38	.29		
6: filler changes timbre ($N = 70$)										
test timbre same	.61	.52	.83	.74	.80	.70	.62	.63	.28	.29
test timbre different	.56	.55	.65	.62	.69	.64	.61	.59	.45	.46
7: different lures copy targets but with contour of melody changed ($N = 31$)	.66	.75	.72	.70	.66	.61	.30	.23	.44	.30
8: similar lures copy targets with pitch level of melody changed ($N = 40$)										
untrained	.64	.61			.61	.61	.41	.41		
moderately trained	.66	.75			.62	.65	.40	.25		

Note: The “filler” material or other conditions of the experiment are noted. Column labels “Short” and “Long” refer to experiment delays of 4s and 12s respectively, except for Experiment 3 and the experiments of Dowling et al. (2002) where they refer to 5s and 15s.

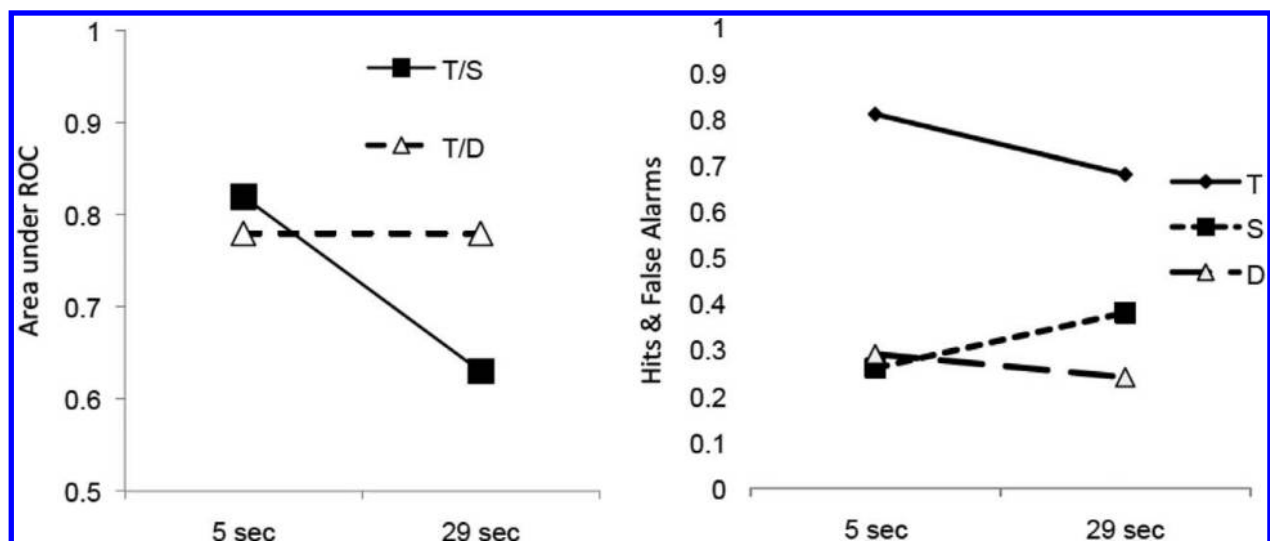


FIGURE 1. Results of Tillmann and Dowling (2007) Experiment 3. (Left) Area under the ROC showing unbiased proportion correct discrimination of target test items from similar lures (T/S) and from different lures (T/D) at short and long delays. (Right) Proportions of hits to target test items (T), and false alarms to similar lures (S) and different lures (D) at short and long delays.

the process of encoding novel music. Shifts in false-alarm rates, such as those found by Dowling et al. (2002), could be due to encoding processes, maintenance of the memory trace, or retrieval processes. In

Experiment 7 below, we report results that lead us to reject a plausible retrieval-based hypothesis for the effect. We believe that encoding processes are responsible for the effect, and so for now we shall focus on them.

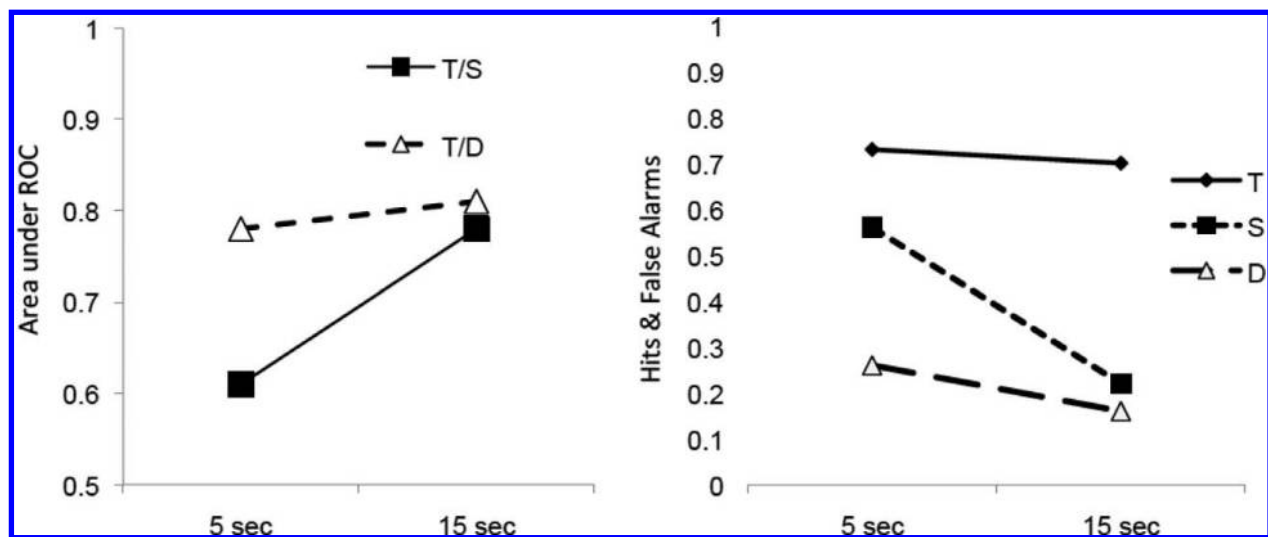


FIGURE 2. Results of Dowling et al. (2002) Experiment 1. (Left) Area under the ROC showing unbiased proportion correct discrimination of target test items from similar lures (T/S) and from different lures (T/D) at short and long delays. (Right) Proportions of hits to target test items (T), and false alarms to similar lures (S) and different lures (D) at short and long delays.

When considered from the perspective of encoding, this result appears to imply that the encoding of musical phrases continues for some seconds after the phrase has been presented, and that it continues while the listener is ostensibly following the ongoing music and beginning the encoding of subsequent phrases. To take a specific example of the kind of continued encoding we have in mind, consider the features of the initial phrases of Beethoven's Minuet in G in Figure 3. In the experiments of Dowling et al. (2002), the initial phrase, designated by bracket 1, could serve as a target (exposition) to be tested after a short delay of two intervening measures with the phrase designated by bracket 3. The latter phrase is a similar lure, sharing the melodic and rhythmic contour of the initial phrase (the pattern of ups and downs of pitch and of relative note durations), but with the melody shifted in pitch along the scale. The melody in bracket 1 begins on the third scale degree and goes up to the fifth, whereas the melody in bracket 3 begins on the first scale degree (the tonic) and goes up to the third. This relationship of the target to the similar lure, in which the melodic contour is shifted along the scale, is typical of the targets and similar lures in those experiments, including the ones of the present study. In their Experiment 1, Dowling et al. (2002) found that after such a short delay (5 s) the false-alarm rate to similar lures was .56, whereas after a longer delay of 6 intervening measures (15 s), the false-alarm rate dropped to .22 (see Table 1). Since the hit rate to target test items (that copied the target) remained relatively constant across this delay (.73-.70), the decrease in false alarms led to an

improvement in area under the ROC from .61 to .78. That is, at the short delay listeners were somewhat confused between target (test) items and similar lures, but after an additional 10 s they were much better able to discriminate them. To explain this, we hypothesize that additional encoding of the original target stimulus takes place during the time between the short and the long delay. We propose that at the short delay, the test item is evaluated just in terms of its individual features. As a similar lure shares the highly salient features of contour and key (musical scale) with the target, the listener is inclined to make a positive response—a false alarm—on the basis of these individual features. During the additional delay, the memory system is able to bind the individual features together, so that not only are the contour and the scale encoded, but also the relationship between them: the place along the scale where the contour is fixed, and thus taking into consideration the tonal functions of the notes. Once the contour is bound to the scale in the memory representation, it becomes apparent that the contour falls on a different scale degree in the similar lure than in the original target, and the test item can be rejected (the correct response).³

³ Based on converging evidence from experiments over the past three decades we believe that recognition of novel melodies in short-term paradigms such as those employed here is primarily based on familiarity, rather than recollection (Dowling, 2010; Dowling et al., 1995). At a short delay similar lures seem familiar because of feature matches with targets. Later, after a longer delay, when a more detailed representation of the target has been formed, mismatches with that detailed representation make the similar lure seem less familiar.

FIGURE 3. Examples of stimuli. The brackets indicate placement of possible target and test phrases in Beethoven's Minuet in G. The $||$ sign indicates an optional repeat of the material preceding this sign before continuing with the remaining part of the minuet (bracket 5). Bracket 1 indicates a test phrase that could be tested after a short delay at bracket 3 with an S. The phrase at bracket 2 could be tested after a long delay with a T (a copy of the original phrase) by taking the repeat. Bracket 5 indicates a potential test with a D of the phrase at bracket 1 that cannot be used because of the intervening imitation at bracket 3. Bracket 4 indicates a potential test with a D after a short delay of bracket 2. The phrase at bracket 6 can test bracket 2 after a long delay with a D in case the repeat is not taken. This example excerpt as well as example excerpts from Experiments 1 to 8 can be found at <http://olfac.univ-lyon1.fr/bt-sound.html>.

This binding hypothesis explains the improvement in false-alarm rates for similar lures over time, and is also consistent with the results of Dowling et al.'s (2002) Experiment 3 (see Table 1). In that experiment, a silent interval (short or long) followed the presentation of the target (exposition), and false-alarm rates to similar lures at the short delay were 20 percentage points less than in their Experiment 1. This suggests that when the listeners are not presented with continuing music to encode between target and test, and hence not distracted from encoding the target, that encoding, including the binding of individual features, proceeds relatively quickly. Overall performance remains strong across both short and long delays, with area scores of .87 and .86. Note that here there is an apparent shift of response criterion between the short and the long delay, in that hits and both types of false alarm decline over the delay, while the area score remains constant. Hence, it appears that the binding of features such as contour and scale is slowed when the listeners are attempting to follow and encode ongoing music at the same time as they are completing the encoding of earlier phrases.

We propose that the binding process just described, along with other encoding processes, is involved in musical structure building, similar to the structure building described by Gernsbacher (1990, 1997) for stories. In musical structure building, a coherent mental representation of a section of a piece is gradually built up during listening by adding information and connections among features as the piece continues. The binding of the melodic contour and a specific scale degree was suggested by Dowling's (1978) theory that a melody is formed cognitively by attaching a contour to a scale. Binding contributes to structure building by transforming the abstract contour into a "dynamic shape" (Jones, Summerell, & Marshburn, 1987) in which the tonal tendencies of the pitches of the melody are brought into play by attaching the contour to the scale, so that differences of stability among the pitches appear in relation to the tonal hierarchy (Krumhansl, 1990), especially in the final pitches of a phrase. As an example in which this change due to binding is stronger than in the preceding example, consider the phrase in bracket 2 (Figure 3). The contour is attached to the scale so that the phrase

ends on the third scale degree. Contrast this with the same contour in bracket 7 at the end of the piece, where the phrase ends on the first degree (tonic), achieving a much stronger sense of closure. The difference in degree of closure achieved by the phrase contributes to the listener's understanding of the structure of the piece—namely, that the phrase in bracket 7 comes at the end of a major section. The overall structure of the piece includes a beginning, a middle, and an end, and these differences in closure, produced by the specific binding of contour and scale, are cues to where a particular phrase falls in that scheme.

There is considerable evidence that listeners use cues such as a cadence, in which the melody ends on the tonic (bracket 7 vs. bracket 2), to decide whether they are in the middle or at the end of a piece. Tillmann, Bigand, and Madurell (1998) presented listeners with halves of 16-bar minuets (similar to those of the present study) and had them decide which were final sections (vs. initial sections). Generally listeners chose as endings the sections that closed with an “authentic” cadence; that is, a cadence in which the harmony progresses from the dominant (V) chord to the tonic (I) chord and the melody ends on the tonic, as in bracket 7. In their experiment, listeners—including musicians and non-musicians—were confused by overall key relationships in the pieces; that is, in pieces in which the initial section changed key but ended with an authentic cadence in the new key (as in bracket 4 of Figure 3), they often chose such an initial section as a final one. This shows that they were still using the closure of the authentic cadence (on the local level) as a cue in their structure building, even if they had difficulties in tracking the global key relationships that placed it in relation to the initial key of the piece.

When interpreting the improvement in performance between the short and long delays as the result of the listeners' continuing structure building while the music continues, we faced the question concerning the conditions under which those processes tend to continue or be interrupted. For stories, Gernsbacher (1990, 1997) found that a structure is developed until a point of closure occurs (such as a scene change in a story). At that point, the structure is summarized as a “gist,” losing details, and a new structure is begun. After replicating the basic result of Dowling et al. (2002) in Experiment 1 here, Experiments 2 through 6 aimed to disrupt the process of continued encoding and structure building using various manipulations of the “filler material” between the target and the test (see Table 1). We found little disruption of the improvement in performance when we scrambled the contents within individual

measures of the filler material by randomizing the order of the beats (Experiment 2), or when we transposed the filler material up one semitone (Experiment 3). However, when we substituted passages from gavottes (in a different key and meter—4/4 rather than 3/4—Experiment 4), or material from other minuets (different key but same meter—Experiment 5), or when the filler changed timbre (Experiment 6), discrimination between targets and similar lures no longer improved. This leads us to conclude that if the ongoing material is perceived as coming from a new piece, or a new source (in the sense of Bregman's, 1990, auditory scene analysis), then structure building is disrupted and the memory system discontinues its encoding of earlier material and discards the previous structure, perhaps starting a new one, as with the stories.

An alternative explanation for our effect, one that emphasizes retrieval rather than encoding, could be raised. This explanation supposes that what changes with increasing delay after the presentation of the target (exposition) is not the state of the listeners' encodings of the target, but rather the availability of melodic contour as a retrieval cue. Dowling and Bartlett (1981), as well as DeWitt and Crowder (1986), suggest that after brief delays, similar lures are confused with target (test) items because of the overwhelming availability of contour as a cue; however, after a longer delay the salience of contour declines, making it possible to compare the test item with the memory for the target (exposition) in terms of other shared and not-shared cues. We tested this alternative explanation in Experiment 7, in which we constructed different lures that differed from target (exposition) items *only* in melodic contour (unlike the target (test)/different-lure pair in Figure 3, brackets 2 and 4, which differ in numerous other features as well). If contour decreased in salience over the longer delay, then discrimination between target (test) items and these different contour lures should decline, and false alarms to those different contour lures (which differed only in contour) should increase. However, the results did not show that pattern, and discrimination between target test items and different-contour lures remained strong after the long delay, supporting our explanation in terms of continued encoding and structure building.

Finally, in Experiment 8, we tested a specific hypothesis generated by our proposed binding framework. If the binding of a melodic contour to the musical scale is the critical accomplishment of the memory system between the short and the long delay, then rejection of similar lures should improve with delay even when a shift of the contour along the scale is the *only* aspect differentiating target (test) items from similar lures.

Because in the previous experiments we were aiming for the ecological validity of an actual listening situation by using the minuets just as written by the composer, similar lures had involved mainly a shift of the melody along the scale, but also other changes as well (as can be seen in Figure 3, brackets 1 and 3—note changes in the bass line). In Experiment 8, similar lures were *exact* copies of targets, except for a shift in the pitch level of the melody. In this case, we replicated the improvement in performance for listeners with moderate amounts of music training, confirming our hypothesis, at least for that group of listeners.

Experiment 1

Experiment 1 was designed as a replication of Dowling et al.'s (2002) Experiment 1 with a new and larger set of stimuli and with delays of 4 and 12 s. The shorter delays came from the use of a faster tempo of 90 beats/min throughout, which made the minuets sound more natural than the previous tempo of 72 beats/min.

METHOD

We provide a detailed method section here, and for subsequent experiments simply note differences. The new set of stimuli used in Experiment 1 was drawn from a different set of minuets than those of Dowling et al.'s (2002) Experiment 1, and so in some of the succeeding experiments we combined the two sets of minuets, giving us a larger set of stimuli to use in creating counter-balanced lists as well as for testing particular comparisons between conditions.

Participants. Forty-four undergraduates at the University of Texas at Dallas (mean age 24.8 yr) served in Experiment 1 as part of their course requirements in psychology. Twenty were categorized as moderately trained in music and had at least 2 yr of explicit music training (defined as lessons on an instrument or voice, or playing in an instrumental ensemble; $M = 6.1$ yr, $SD = 4.1$ yr). Twenty-four participants with less training were categorized as musically untrained. Since music training was involved in neither simple effects nor interactions in the analyses of any of the experiments except Experiment 8, we have omitted it from the analyses reported for the first seven experiments. No participant served in more than one experiment. None indicated that they had absolute pitch, though we did not ask explicitly about that.

Stimuli. The stimuli were drawn from classical minuets, waltzes, German dances, and Ländler written for piano between 1750 and 1828 by Haydn (1984, 1989), Mozart

(1956, 1992), Beethoven (1967, 1987, 1990), and Schubert (1989—see also Lakos, 1994, for a piece by Handel). We included some minuets from Mozart's sonatas for violin and piano, adding the violin notes to the piano part. These dances in 3/4 time followed a form in which an initial section of 8–12 measures (delimited by a repeat sign) was followed by a (usually longer) section of 8–32 measures (see Figure 3 for an example, but one that we used only as a sample stimulus in the experiments because of its widespread familiarity). In the large music library at Southern Methodist University, we found 106 of these minuets that met our criteria, drawn from among all the minuets in the volumes just cited plus a few miscellaneous collections of similar pieces, such as Lakos (1994). We selected pieces in which potential target phrases were neither repeated nor imitated in the continuation of the piece prior to potential test phrases. The test phrases were repetitions of the target phrase (as in Figure 3, bracket 1), imitations that preserved the melodic and rhythmic contour of the target phrase (similar lures, Figure 3, bracket 3), or phrases differing in rhythmic and melodic contour from any preceding phrase (different lures, Figure 3, bracket 6), all just as the composer had written them. We made a different random selection of these pieces for each of the following experiments, with the constraint that equal numbers of each trial type would be represented (since not all minuets afforded all of the six types of trial: tests of targets, and similar and different lures at short and long delays).

On each trial, participants heard the first 12 to 20 s of a piece. One of the first two phrases of the piece (during the first 8 s, or roughly 4 measures) was chosen as a target (exposition) to be tested later. The music was presented without any indication about where the target phrase was situated. Half of the targets consisted of the first phrase of the piece (roughly, measures 1 and 2) and half the second phrase (roughly, measures 3 and 4). A new piece appeared on every trial. The minuets were presented at a tempo of 90 beats/min, so that each 3/4 measure took 1 s. Hence test phrases occurred within the piece after delays following the presentation of the target of 4 or 12 s—that is, with 2 or 6 measures of intervening “filler” material, which was exactly what the composer had written. Since the test phrases all occurred exactly as they had in the original score, whatever contextual cues there were to their content applied equally for all three test item types.

The onset of the test was signaled by a soft, high-pitched “beep,” separated from the rest of the music by at least two octaves, and occurring one-half beat before the test (that is, on an unoccupied off-beat, and

so conflicting with no other notes). The music stopped after presentation of the test phrase and participants were given 10 s to respond before the beginning of the next trial.

The test phrase was either a target (test) —an exact repetition of the target (exposition) —or an imitation of the original target that changed one or more features of the target but left the melodic and rhythmic contour largely intact (similar lure), or a new phrase not heard before in the piece (different lure). For example, the phrase in bracket 2 of Figure 3 could be tested as a target (test) after a long delay by taking the repeat at the || sign and coming back to the same phrase after 6 intervening measures. The phrase in bracket 1 could be tested as a similar lure after a short delay (two measures) with the phrase in bracket 3. And the phrase in bracket 2 could be tested with a similar lure after a long delay with the phrase in bracket 6. (Note that using a target (test) did not always involve a repeat; this is just a characteristic of the present example, which has an eight-bar first section.) Note that the phrase in bracket 1 could not be tested after a long delay with the different phrase in bracket 5 because the phrase in bracket 1 is imitated in bracket 3, which is likely to remind the listener of the target (exposition) phrase.

There were six conditions defined by the combination of two delays (4 s or 12 s) and three types of test phrases (target (test), similar lure, or different lure). There were 48 trials in all, with eight of each type. The order of trials was randomized so that the list consisted of four permutations of blocks of trials containing twelve trials. A block consisted of six trial types each involving target (exposition) items occurring in either measures 1-2 or 3-4. This ensured that each type of trial would be tested equally often during various phases of the list. The list of trials was divided into two equal sections, and half the participants heard the sections in reverse order. This was done to minimize carry-over effects.

Stimuli were played by the first author, one hand at a time and at a slow tempo (60 beat/min), on a Yamaha Clavinova P-100 (which has weighted piano keys and touch-sensitive responses) and recorded by a PC-type computer via its MIDI interface. Though no attempt was made to make these performances “expressive” in a broader sense, particular attention was given to articulation, phrasing, and dynamics within each phrase to make the performance as natural and aesthetically pleasing as possible within these constraints. The techniques used were essentially those taught in the first 2 yr of piano study. We used Cakewalk software to edit the recordings, correct recording errors, combine the voices, and impose a uniform tempo. The tempo was

a uniform 90 beats/min to produce short and long delays between the introduction of an item and its test of 4 and 12 s (except where noted). Even though local tempo variations within each piece would have sounded more natural, such variations are context dependent—determined by the place of a phrase within a section (Gabrielsson, 1999).⁴ We used the Cakewalk editor to insert the exact repetitions of targets that served as target (test) items. Stimuli were stored as midi files and played for participants on a computer controlled Yamaha TG-500 synthesizer using its “acoustic piano” voice and presented to participants in group sessions of three to five listeners via loudspeakers at comfortable levels (except in Experiments 7 and 8 where they were recorded as CD-quality .wav files and presented via headphones). Listeners with different amounts of training were not assigned to different group sessions.

Procedure. Participants were introduced to the experiment by a brief explanation of the task including examples of the differences among target (test) items, similar lures, and different lures. The instructions emphasized that participants should respond *same* only when the test item was *exactly* the same as the target, and to reject similar lures as well as different lures; earlier research indicates that participants would be likely to find it difficult to do otherwise (Dowling & Bartlett, 1981). Participants were instructed to respond with pencil and paper using a six-point confidence-level scale on which 6 = *very sure same*, 5 = *sure same*, 4 = *same*, 3 = *different*, 2 = *sure different*, and 1 = *very sure different*. In Experiments 7 and 8 participants responded with number keys on the computer. Participants also completed a brief questionnaire concerning music training.

Data analysis. Responses to the three test item comparisons were reduced to two areas under the ROC (receiver operating characteristic), one assessing discrimination between target (test) items and similar

⁴It is true that the dynamics (loudness pattern) of a phrase can be affected by context, thus making similar lures occurring at the ends of sections less similar to targets than similar lures in the middle of a section (and hence likely to be at a shorter delay). However, we doubt that this had an effect on the results. A new experiment (Tillmann et al., 2013) utilized stimuli played expressively by an expert pianist who used a wide range of tempos and introduced expressive differences between target (exposition) items and test items. The results were qualitatively similar to those reported here. And in Experiments 7 and 8 we used different lures that were exact copies of targets except for changes in pitch contour (Experiment 7), and similar lures that were exact copies of targets except for the pitch level of the melody (Experiment 8).

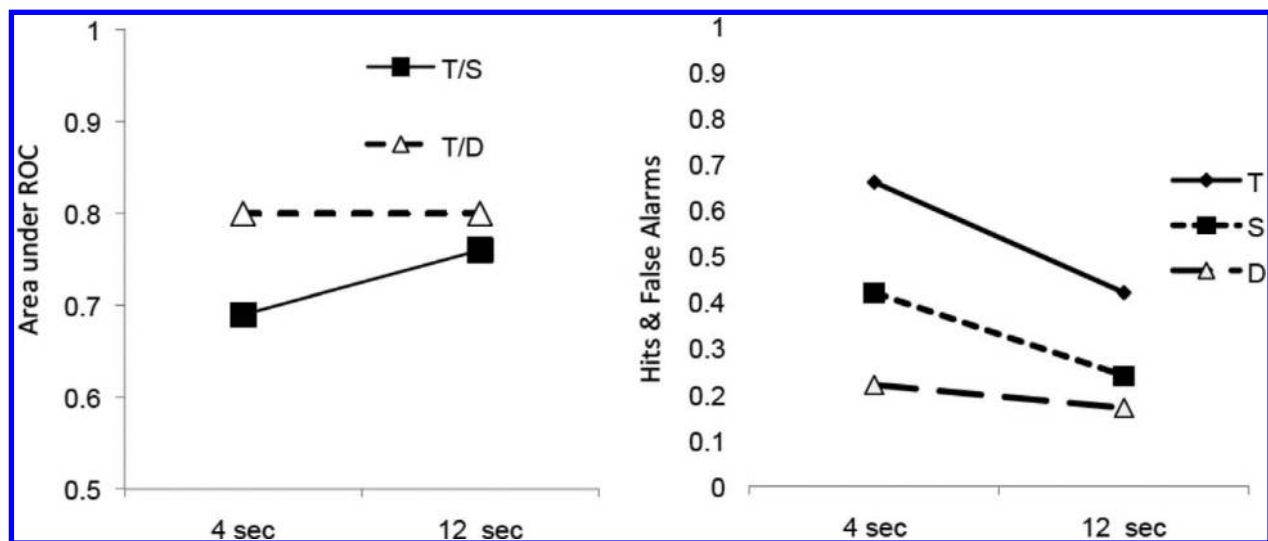


FIGURE 4. Results of Experiment 1, a replication of Dowling et al. (2002) Experiment 1. (Left) Area under the ROC showing unbiased proportion correct discrimination of target test items from similar lures (T/S) and from different lures (T/D) at short and long delays. (Right) Proportions of hits to target test items (T), and false alarms to similar lures (S) and different lures (D) at short and long delays.

lures, and the other discrimination between target (test) items and different lures (referred to as T/S and T/D discrimination in Table 1 and the figures). Each area score was based on 16 data points: 8 hit rates and 8 false alarm rates. The six-point scale provided us with five criterion placements on the ROC with which to calculate the area. Area under the ROC provides an unbiased estimate of proportion correct where chance is .50 (Swets, 1973). The area score provides a better measure of performance than, for example, d' based on the criterion between responses 3 and 4, because it preserves more response information and over the years has been seen to be uncorrelated with measures of bias (unlike d' —see Dowling, Kwak, & Andrews, 1995; Verde, Macmillan, & Rotello, 2006).

We report analyses of variance (ANOVAs) of the area scores in Table 1, along with the proportions of hits and false alarms to target (test) items and the two kinds of lure. We report the results of t -tests on the differences for those proportions between the short and the long delays. What is important to our argument is first, to see whether the area scores assessing discrimination between target (test) items and similar lures increase with delay, and then to see if that improvement is accompanied by a significant decrease in false alarms to similar lures. Without the improvement in the bias-free area score, changes in the similar lure false alarm rate might merely reflect changes in response criterion when seen in conjunction with decreases in hits and similar-lure false alarms. Inspection of Table 1 will reveal cases in which

hit and false-alarm rates change without a change in area scores. This was observed in the present Experiment 4 and in Dowling et al.'s, (2002) Experiment 3, in which all three proportions of hits and false alarms declined over time, but with little effect on area scores for target (test) items and similar lures.

RESULTS & DISCUSSION

The area scores from Experiment 1 (Table 1 and Figure 4A) were subjected to a 2 Delay (short/long) X 2 Item Comparison (target (test) vs. similar lures/target (test) vs. different lures) repeated measures analysis of variance (ANOVA). Discrimination between target (test) items and different lures was better than that between target (test) items and similar lures, $F(1, 43) = 27.70$, $p < .001$, $MSE = 0.008$. The interaction between Delay and Item Comparison, $F(1, 43) = 7.44$, $p < .01$, $MSE = 0.007$, indicated that performance improved with delay for discrimination between target (test) items and similar lures, $t(42) = 2.28$, $p < .05$, but not for discrimination between target (test) items and different lures. No other effects were significant.

The improvement in discrimination between target (test) items and similar lures was accompanied by a significant decrease in similar lure false alarms, from .42 to .24, $t(42) = 3.84$, $p < .001$ (see Table 1 and Figure 4B). There was also an additional criterion shift leading to significant declines in hits, $t(42) = 2.76$, $p < .01$, and in different-lure false alarms, $t(42) = 2.47$, $p < .01$. Note that although the area scores in Experiment 1

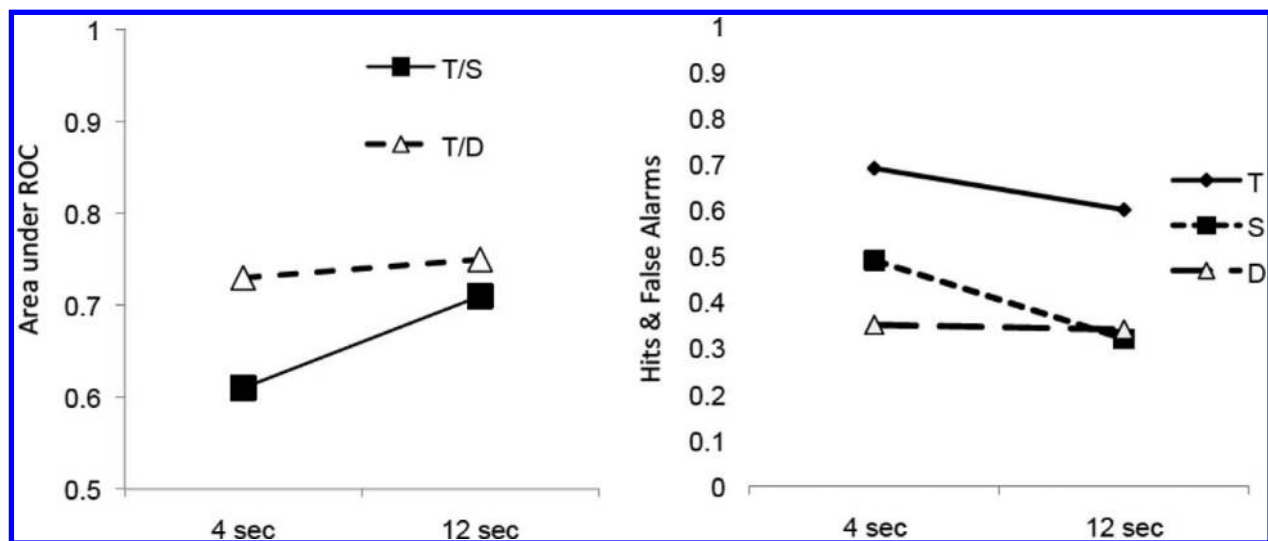


FIGURE 5. Results of Experiment 2, in which the order of beats in the filler material was scrambled within measures at short and long delays. (Left) Area under the ROC showing unbiased proportion correct discrimination of target test items from similar lures (T/S) and from different lures (T/D). (Right) Proportions of hits to target test items (T), and false alarms to similar lures (S) and different lures (D) at short and long delays.

were in the same range as those of Dowling et al.'s (2002) Experiment 1, the hit rates were generally lower, perhaps due to the different selection of stimulus items.

Experiment 2

The results of Experiment 1 replicated those of Dowling et al. (2002, Experiment 1) with a new set of musical pieces. Having replicated those results, we wanted to explore the limits of this phenomenon. We thought that it might be the case that the coherence of the "filler" material occurring between target (exposition) and test might be important to the improved rejection of similar lures with greater delay. Hence we decided to increase the complexity and reduce the coherence of the filler material while preserving its overall structure, to see if this would produce increased interference and disrupt the continued encoding of the initial phrases. Therefore in Experiment 2 we scrambled the contents of measures in the filler. We randomly rearranged the order of the beats within each measure, so that everything contained in the first beat was moved to beat 2 or beat 3, etc. This destroyed the details of the melodic line and the harmony, while preserving both in broad outline.

METHOD

Participants. Twenty-six participants served in Experiment 2; 11 were moderately trained and 15 untrained.

Stimuli. Stimuli consisted of a new sample of 48 minuets from the 106 available, and were constructed as in Experiment 1, with the contents of all of the measures between the end of the target (exposition) and the beginning of the test rearranged randomly. We divided the contents of each measure into three packages corresponding to the three beats, and randomized the order of those packages, with no two successive measures rearranged in the same way.

To assure ourselves that the scrambling produced a deleterious effect on the perceived structure, we ran a control experiment in which we asked ten participants not taking part in any of the experiments to rate the stimuli from both Experiments 1 and 2 for coherence on a ten-point scale. The stimuli with intact music received higher coherence ratings (7.0) than the stimuli with scrambled measures (5.9), $F(1, 9) = 10.61$, $p < .01$.

RESULTS & DISCUSSION

Area scores (Table 1 and Figure 5A) were subjected to a 2 Delay X 2 Item Comparison repeated measures ANOVA. Discrimination between target (test) items and different lures was better than that between target (test) items and similar lures (.74 vs. .66), $F(1, 25) = 11.33$, $p < .01$, $MSE = 0.015$. The Delay X Item Comparison interaction, $F(1, 25) = 5.20$, $p < .05$, $MSE = 0.009$, showed that area scores increased with delay for discrimination between target (test) items and similar lures (from .61 to .71), $t(24) = 2.35$, $p < .05$, but not for discrimination between

target (test) items and different lures. No other effects were significant.

Proportions of false alarms to similar lures declined with delay from .49 to .32, $t(24) = 12.68$, $p < .001$. Again, there was a criterion shift, in which hits declined by .09, $t(24) = 6.71$, $p < .001$, and different lure false alarms by .11, $t(24) = 8.20$, $p < .001$ (see Table 1 and Figure 5B).

Memory improvement was observed again, even with scrambled measures in the filler. As in Experiment 1, the increase in area scores for target (test) items vs. similar lures across the delay was produced by a drop in false alarms to similar lures, in spite of a concurrent decrease in hit rates. Overall performance was somewhat worse than in Experiment 1, probably due to the distracting nature of the scrambled filler material. The coherence judgments showed that scrambling produced a deleterious effect on perceived structure. If some level of structural integrity is necessary to link the original target (exposition) and the test in order to produce memory improvement, the necessary linkage appears to operate at a broader level than within-measure details, since destroying some of those details did not eliminate the discrimination improvement effect.

Experiment 3

We were surprised that scrambling the contents of measures in Experiment 2 did not disrupt the improvement of target (test) vs. similar lure discrimination observed in Experiment 1. We wondered if shifting the key of the filler material by one semitone would cause a disruption in continued processing. Therefore in Experiment 3 we kept the original continuation of the minuet as in Experiment 1, but changed the key of the entire filler material by transposing it up one semitone, shifting back to the key of the original target (exposition) for the test item. A shift by one semitone does not disrupt the melodic contour or rhythmic continuity; however, it does involve a modulation to a distant key, thus introducing some tonal incoherence and a shift to a largely new set of pitches.

METHOD

Participants. Thirty-seven participants served in Experiment 3; 21 were moderately trained and 16 untrained.

Stimuli. Stimuli consisted of a new sample of minuets selected from the pool of available minuets, and were constructed as in Experiment 1, but with the material between the end of the target (exposition) and the beginning of the test raised in pitch by one semitone. The stimuli were presented at a tempo of 72 beats/min,

resulting in delays of 5 and 15 s, because this experiment was done as part of the series of experiments described in Dowling et al. (2002), before we introduced the faster tempo of 90 beats/min.

RESULTS & DISCUSSION

Area scores (Table 1 and Figure 6A) were subjected to a 2 Delay X 2 Item Comparison repeated measures ANOVA. There was better performance at the long delay (.78) than the short delay (.70), $F(1, 36) = 9.91$, $p < .01$, $MSE = 0.024$. Performance with comparisons between target (test) items and different lures was better (.79) than with comparisons between target (test) items and same lures (.69), $F(1, 36) = 63.47$, $p < .001$, $MSE = 0.006$. The interaction between Delay and Item Comparison, $F(1, 36) = 14.76$, $p < .001$, $MSE = 0.005$, indicated improvement with delay for comparisons between target (test) items and similar lures (from .63 to .75), $t(35) = 4.05$, $p < .001$, but not for comparisons between target (test) items and different lures (*ns*). No other effects were significant.

Proportions of false alarms to similar lures declined sharply with delay from .51 to .27, $t(35) = 4.92$, $p < .001$, whereas hits remained constant and false alarms to different lures declined, but not significantly (see Table 1 and Figure 6B).

The shift in key of the filler material had little effect on discrimination improvement, and overall performance was about the same as in Experiment 1. It may be that the shift upward by one semitone was hardly noticed by our moderately trained and untrained participants, since such a shift is common in a wide range of popular music from country and Western to Christmas songs, where successive stanzas of a song are often presented one semitone higher to add excitement. And although a half-step transposition represents quite a distant modulation (five steps around the circle of fifths), musically untrained participants might be more attuned to pitch-height differences than to distances in terms of the tonal framework (Krumhansl & Shepard, 1979). Furthermore, the return to the original key, down a half step, simply put the test item in the same key as the target (exposition). Previous research has shown that for both musician and nonmusician listeners, incoherence in tonality does not influence target detection in short minuets. Tillmann and Bigand (1998) found that memory performance was equally good for minuets presented coherently (with respect to tonality) as for minuets with parts transposed to different keys that were close in pitch height but tonally distant in terms of the circle of fifths.

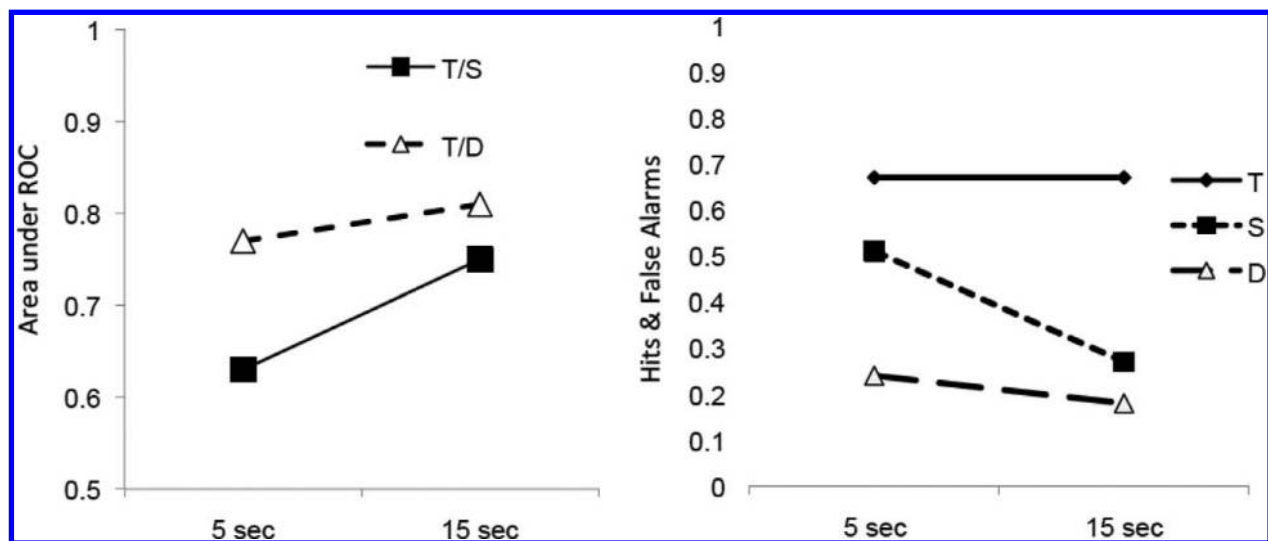


FIGURE 6. Results of Experiment 3, in which the filler material was transposed up 1 semitone. (Left) Area under the ROC showing unbiased proportion correct discrimination of target test items from similar lures (T/S) and from different lures (T/D) at short and long delays. (Right) Proportions of hits to target test items (T), and false alarms to similar lures (S) and different lures (D) at short and long delays.

Experiment 4

In Experiment 2 and 3, neither reducing the coherence of the filler material by scrambling within measures, nor shifting the filler by one semitone had much effect on memory performance. In Experiment 4 we produced what we thought might be greater incoherence by substituting passages from gavottes for the continuations of the minuets. Gavottes have a duple meter (being in 4/4 time) rather than a triple meter like the minuets, so that their substitution results in considerable metrical (as well as thematic) incoherence. We also introduced the gavotte material in a different but closely related key to that of the minuet, an additional difference we thought should produce incoherence.

METHOD

Participants. Twenty participants served in Experiment 4; 10 were moderately trained and 10 untrained.

Stimuli. Stimuli consisted of a new sample of 48 minuets from the total set of 106, and were constructed as in Experiment 1, but with excerpts from gavottes drawn from the same historical/stylistic period substituted for the material between the end of the target (exposition) and the beginning of the test, replacing the same number of beats (the same time delay) as had been there in the minuets. The passages from the gavottes were drawn from the same composers' albums of dances as well as from Lakos (1994), utilizing material that was structurally parallel to the to-be-replaced minuet passages. That

is, they typically involved phrases beginning around the third or fifth measure of the first section of the gavotte. The initial beat of a gavotte phrase (on the third beat of a 4/4 measure) was aligned with the initial beat of the minuet phrase following the target (exposition) presentation, and the interpolated gavotte material replaced the minuet material and continued for the requisite number of beats at the same tempo. We minimized disruption of the overall contour and texture of the minuet by choosing gavotte passages that joined the adjacent minuet with relatively smooth voice leading in all the parts. The gavottes were in nearby keys to the minuets (one or two added flats or sharps; that is, one or two steps around the circle of fifths). Because of the possibility that the two sets of minuets we were working with might be differentially memorable (cf. the disparity in hit rates in Table 1) we made two separate lists for each condition, one using the stimulus set from Experiment 1 and the other the stimulus set from Experiment 1 of Dowling et al. (2002). Approximately equal numbers of participants performed the experiment with each list. There were no significant effects arising from the difference between lists. There were 48 trials in each list.

RESULTS & DISCUSSION

Area scores (Table 1 and Figure 7A) were subjected to a 2 Delay X 2 Item Comparison repeated measures ANOVA. Performance was better for discrimination between target (test) items and different lures (.79) than for comparisons between target (test) items and similar

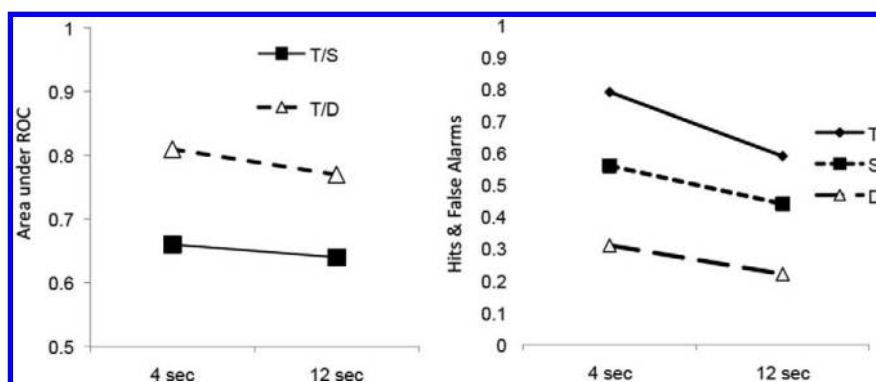


FIGURE 7. Results of Experiment 4, in which a gavotte was substituted in the filler material. (Left) Area under the ROC showing unbiased proportion correct discrimination of target test items from similar lures (T/S) and from different lures (T/D) at short and long delays. (Right) Proportions of hits to target test items (T), and false alarms to similar lures (S) and different lures (D) at short and long delays.

lures (.65), $F(1, 19) = 68.63$, $p < .01$, $MSE = 0.006$. No other effects were significant.

The proportions of similar-lure false alarms, $t(18) = 2.85$, $p < .01$, and hits, $t(18) = 3.01$, $p < .01$, declined with delay, and different-lure false alarms declined by a non-significant amount, due to an apparent shift toward more conservative response criteria (fewer “same” responses) at the longer delay (see Table 1 and Figure 7B).

Switching to gavottes with different meter and different key during the delay apparently produces sufficient incoherence to eliminate the discrimination improvement effect. Note that even though similar-lure false alarms declined by 12 percentage points, the accompanying hit rates declined by 20 percentage points and the area score for discrimination between target (test) items and different-lure false alarms remained virtually constant; there was clearly a shift in response bias from the shorter to the longer delay. It may be that metric-rhythmic continuity is necessary for the discrimination improvement effect. What would happen if we substituted a different minuet, leaving the meter constant but changing the thematic content and key?

Experiment 5

Since a switch to gavottes coupled with changes in meter and key was disruptive of memory improvement, we decided to substitute different minuets in the filler material (that is, same meter but different thematic content and in a different key). In a methodological refinement over Experiment 4, we made the substitution only for the two measures preceding the test. We did this because in Experiment 4, there were 6 beats of disruptive material at the 4-s delay, but 18 beats of disruptive material at the 12-s delay. It is conceivable that the

poorer performance at the longer delay was simply due to the greater amount of disruptive material, and not specifically due to disruption of the melodic-rhythmic link between target (exposition) and test. Therefore in Experiment 5, we kept the amount of different interpolated material constant at two measures in both the short and the long delays. In addition to substituting material from other minuets of the same period (a substitution comparable to that of Experiment 4), we included a control condition in which the minuet simply continued as written—essentially a replication of Experiment 1. We tested only targets and similar lures, since those provide the critical comparisons for the memory-improvement effect previously observed.

METHOD

Participants. Thirty-five participants served in Experiment 5; 17 were moderately trained and 18 untrained. Approximately equal numbers of participants at each level of training were assigned blindly to conditions in which the different minuet was in a near key (one or two additional flats or sharps) or a far key (four or five additional flats or sharps).

Stimuli. Stimuli consisted of a new sample of 48 minuets of the total of 106 available, and were constructed as in Experiment 4, but using only the 3-4 target position and only target (test) items and similar lures as test items for short and long delays. Experiment 5 introduced an additional within-participant variable; type of interfering material (same vs. different minuet). The same-minuet condition was a replication of Experiment 1. In the different-minuet condition the two measures preceding the test item were replaced with two measures from the corresponding passage in a different minuet

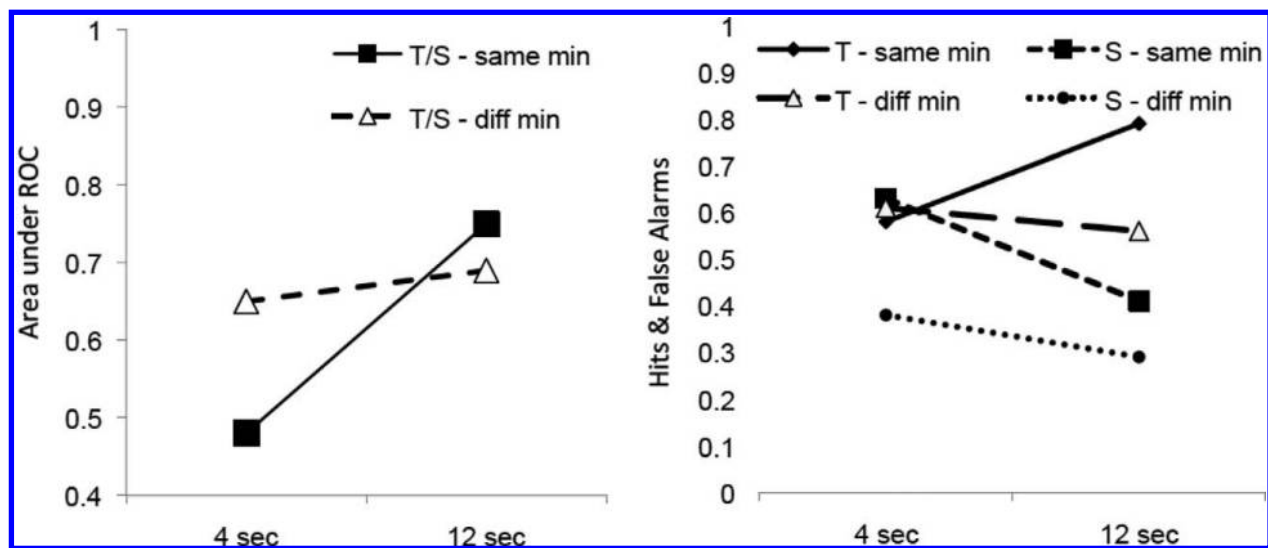


FIGURE 8. Results of Experiment 5, in which the last two measures of the filler material were replaced with a different minuet, or not. (Left) Area under the ROC showing unbiased proportion correct discrimination of target test items from similar lures (T/S) for same minuet and different minuet filler at short and long delays. (Right) Proportions of hits to target test items (T), and false alarms to similar lures (S) for same minuet and different minuet filler at short and long delays.

that did not otherwise appear in the experiment. The material from the different minuet was always in a different key from that of the original: half were in a key close to that of the original (one or two added flats or sharps as in Experiment 4) and the other half were in distant keys (four or five added flats or sharps). An analysis of the data revealed that there was no effect of key distance of the interpolated material, so we collapsed items across key distance for the analyses reported here. In sum, there were 24 trials in each of the two conditions (same vs. different minuet), providing 6 trials with each type of test item.

RESULTS & DISCUSSION

Area scores (Table 1 and Figure 8A) for discrimination between target (test) items and similar lures were subjected to a 2 Minuet (same vs. different) X 2 Delay repeated measures ANOVA. Performance was better with increasing delay, $F(1, 34) = 24.47, p < .001, MSE = 0.035$. The Minuet X Delay interaction, $F(1, 34) = 13.15, p < .01, MSE = 0.034$, showed that there was stronger improvement with the same minuet (.48 to .75), $t(33) = 3.62, p < .01$, than with the different minuet (.65 to .69), *ns*. Thus, these results replicated Experiment 1 with the continuation of the same minuet, but those results were disrupted with the different minuet. There were no other significant effects.

In the same-minuet condition, similar-lure false alarms decreased from .63 to .41, $t(33) = 3.51, p < .001$, and hits

increased from .58 to .79, $t(33) = 3.51, p < .001$, over time (see Table 1 and Figure 8B). In the different-minuet condition, there were no significant changes in hits and false alarms.

The discrimination improvement effect disappeared when two measures of a different minuet were interpolated before the test. Note that even though similar-lure false alarms declined by 9 percentage points with the different minuet filler, the decline was accompanied by a decline in hit rates of 5 percentage points, leading to only a slight, nonsignificant increase in the area scores; there was obviously a shift to a more cautious response criterion in the longer delay with the different-minuet filler (see Table 1). Just two measures of a different minuet are sufficient to break the continuity linking target (exposition) and test and to disrupt continued encoding of the target phrase. Because the interpolated minuet differed from the target minuet in both key and thematic material, it is not possible to attribute this disruption *solely* to one or the other factor, or to its accumulating effect. But note that solely changing key on its own does not disrupt the effect (Experiment 3),

Experiment 6

In Experiments 2-5, we kept the relationship between targets and similar lures relatively constant while varying the nature of the filler material in the interval between target (exposition) and test. The manipulations

of filler material in Experiments 2 and 3, which focused on local violations of thematic material and on modest changes in harmonic relations, did not disrupt the musical continuity sufficiently to eliminate the discrimination improvement effect observed in Experiment 1. In contrast, the manipulations of Experiments 4 and 5, which radically altered global thematic and key relationships, were disruptive. It seems that introducing material from another piece, and in a noticeably different key, is sufficient to inhibit ongoing memory processing of potential target (exposition) items in the interval before the test.

In general, the features we altered in the filler material in Experiments 2 through 5 were features that remained constant between targets and similar lures. Those features are important for defining the similarity of target (test) items and similar lures: tonality, meter, melodic contour, and rhythm. In addition to those features, there is also timbre, which, in this style, like meter, generally remains constant throughout the piece. In Experiment 6, we changed the timbre of the filler, keeping the same timbre for target (exposition) and test, while preserving thematic and harmonic continuity. If what is disruptive of memory encoding in the shifts to other pieces in Experiments 4 and 5 is an apparent change of sound source (in the sense of auditory scene analysis, Bregman, 1990), then a change of timbre should be at least as disruptive. We also introduced an additional condition: after the change of timbre for the filler material: the timbre of the test item could either return to the timbre of the target (exposition), or shift to yet a third timbre. We thought that returning to the original timbre for the test item would lead to better performance than changing to a third timbre, on the supposition that recognition would be based on a match of sensory features of the target (exposition), such as the original timbre.

METHOD

Participants. Seventy participants served in Experiment 6; 38 were moderately trained and 32 untrained.⁵

Stimuli. Stimuli consisted of a new sample of 96 minuets from the pool of available minuets, and were constructed as in Experiment 1, with the additional within-participant variable of test-item timbre (same as or different than the original target timbre). On half the trials, the test timbre was the same as that of the

original target, and on half the trials it was different, randomly determined. Each trial in the experiment began with the same timbre from the set available on the Yamaha TG-500 synthesizer (for example, “bright acoustic piano”), and after the target (exposition) was presented switched to another timbre (for example, “vibraphone”) for the filler material. The test item was presented in either the timbre of the target (exposition) or a different, third, timbre (for example, “acoustic guitar—steel”). We used these three timbres throughout as they are all similar in their attack and decay times. The assignment of the three timbres to the three stimulus components was systematically permuted in six counterbalanced lists across the 96 trials.

RESULTS & DISCUSSION

Area scores (Table 1 and Figure 9A) were subjected to a 2 Test Timbre X 2 Delay X 2 Item Comparison repeated measures ANOVA. Performance was better for same timbre (.68) than different timbre (.60) test items, $F(1, 69) = 46.05, p < .001, MSE = 0.019$. Performance was better at the shorter (.66) than the longer (.61) delay, $F(1, 69) = 19.40, p < .001, MSE = 0.020$. Performance was better for target (test) vs. different lure discrimination (.71) than for target (test) vs. similar lure comparisons (.56), $F(1, 69) = 173.75, p < .001, MSE = 0.018$. The interaction of Test Timbre X Delay indicated a decline over time for same-timbre test stimuli but not for different timbre, but this was clearly influenced by a floor effect (see Table 1), $F(1, 69) = 6.90, p < .02, MSE = 0.023$. And the interaction of Test Timbre X Item Comparison was similarly influenced by a floor effect for discrimination between target (test) items and similar lures in the different timbre condition, $F(1, 69) = 72.14, p < .001, MSE = 0.009$. There were no other significant effects.

Proportions of hits declined significantly over time with same-timbre tests, $t(68) = 2.82, p < .01$, accompanying the decline in area scores (see Table 1 and Figure 9B). False alarms to similar lures and different lures remained relatively constant over delays in all conditions of the experiment.

Changes of timbre were very disruptive of the continuity that seems necessary for memory improvement. Participants very likely heard the different timbre filler material as coming from a different source, since in Bregman’s (1990) auditory scene analysis timbre is a strong cue for source attribution. This shift of source attribution parallels what happened with the gavottes and different minuets in Experiments 4 and 5, leading to the discontinuation of the encoding and feature binding of target (exposition) items.

⁵The large number of participants in this experiment was due to the fortunate accidents of the availability of a relatively large number of volunteer research assistants and of participants needing research credit.

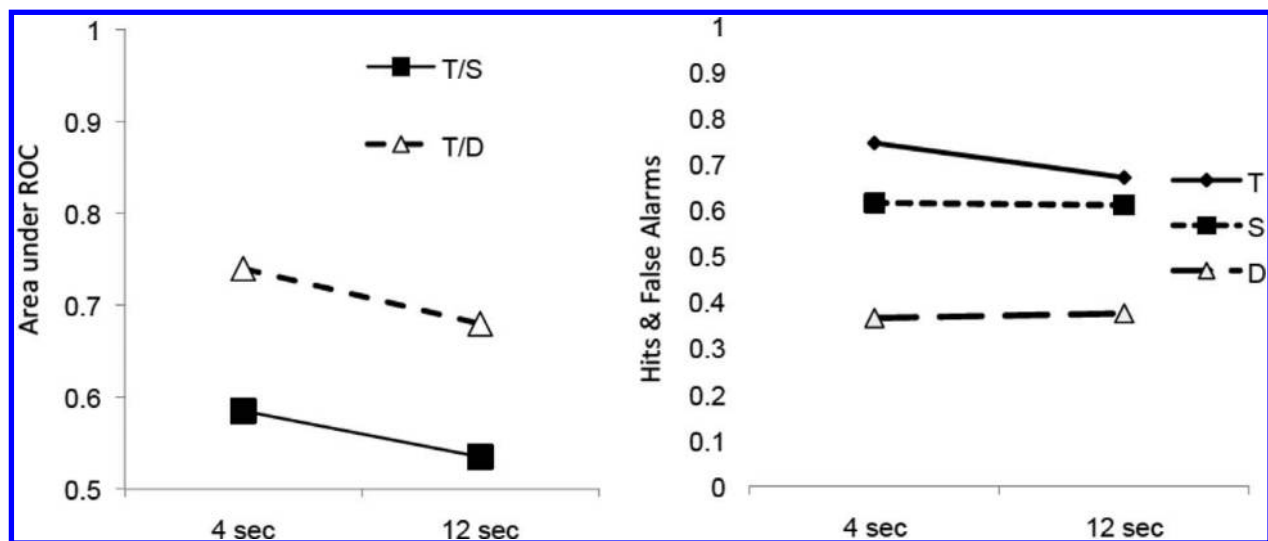


FIGURE 9. Results of Experiment 6, in which the filler material changed timbre. (Left) Area under the ROC showing unbiased proportion correct discrimination of target test items from similar lures (T/S) and from different lures (T/D) at short and long delays. (Right) Proportions of hits to target test items (T), and false alarms to similar lures (S) and different lures (D) at short and long delays.

In summary, Experiments 2 through 6 show that changing melodic organization at a very local level, and changing the key by a semitone, do not disrupt ongoing encoding and structure building between target (exposition) and test, whereas changes that appear to indicate a change of source, either by changing the piece in midstream or changing the timbre, do disrupt continued encoding.

Experiment 7

In Experiment 7, we tested the alternative hypothesis described in the Introduction that emphasizes retrieval conditions rather than encoding as the source of the change in performance between the short and long delays. This hypothesis suggests that the fading salience of contour as a retrieval cue over time is responsible for the relative lack of confusion at the long delay between target (exposition) items and same-contour similar lures. However, if contour loses its salience as a cue, we would also expect performance in discriminating target (test) items from different-contour lures to decline. This has in general not been observed in the previous experiments: in six conditions there was an average decline of 1.7 percentage points in discrimination between target (test) items and different lures, and in only one of those conditions was there a decline of more than 4 percentage points (see Table 1). In addition, the different lures in those experiments differed from

targets in more ways than just a difference of contour. Consider, for example, the target (test)/different-lure pairs in Figure 3 of bracket 2/bracket 4 and bracket 2/bracket 6 (short and long delay tests of the bracket 2 target (exposition), respectively). Numerous features besides simply the melodic contour are altered between the original target and the different lure at test. Therefore, since looking at the previous different lures does not provide the answer, in Experiment 7 we constructed different lures that differed from targets *only* in melodic contour. This was achieved by making an exact copy of the target (exposition), and then altering the melody so that at the third or fourth note the contour changed direction with respect to the original target. For example, the melody in bracket 2 (Figure 3) would leave the first three notes intact, but then alter the contour by moving downward thereafter, preserving the same rhythm and the other aspects of contour and harmonic accompaniment. If contour becomes less important as a recognition cue with longer delays, then in this case discrimination between target (test) items and different lures should decline with increased delay.

METHOD

Participants. Thirty-one participants served in Experiment 7; 15 were moderately trained and 16 untrained.

Stimuli. Stimuli consisted of a new sample of 72 minuets drawn from all those used in Experiments 1 and 2, and were like those used there, except that different lures

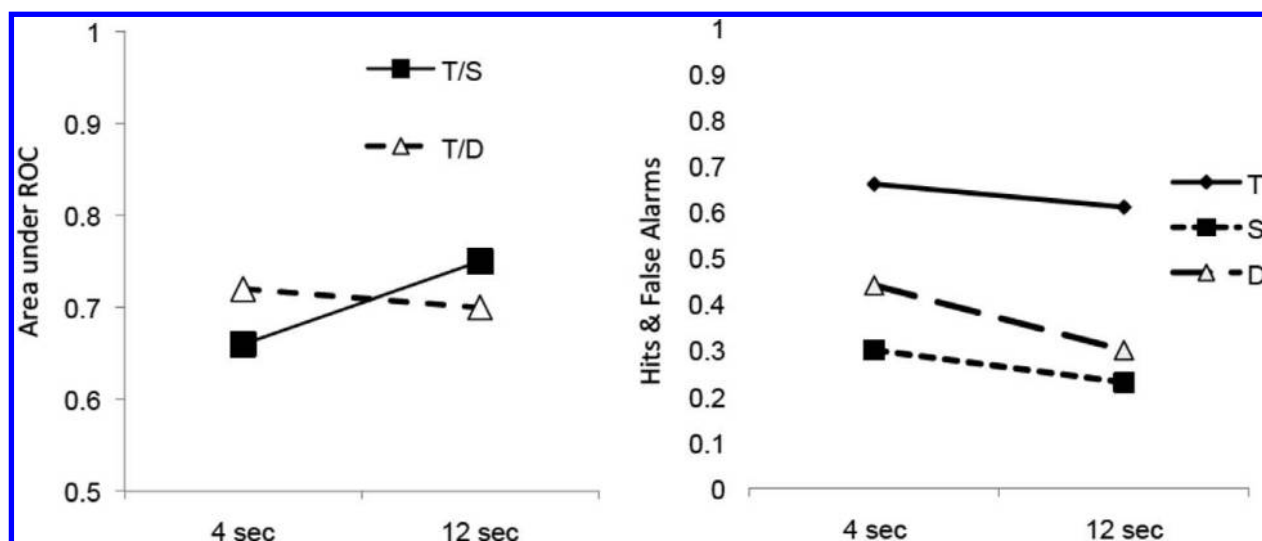


FIGURE 10. Results of Experiment 7, in which different lures were copies of targets in which the melody changed direction about a third of the way through. (Left) Area under the ROC showing unbiased proportion correct discrimination of target test items from similar lures (T/S) and from different lures (T/D) at short and long delays. (Right) Proportions of hits to target test items (T), and false alarms to similar lures (S) and different lures (D) at short and long delays.

were copies of target (exposition) items in all respects *except* the contour of the melody. We used Cakewalk software to copy the target onto the different-contour lure test item, and then altered the pitches of the melody to change its contour in a way that conformed to the harmony of the original target. The changes started with the third or fourth note and generally altered four or five pitches and two or three directions of intervals in the melody.

There were 72 trials, recorded as .wav files and presented by means of a Matlab program (version 2009b) that produced a different random order of trials for each participant. In Experiments 7 and 8, participants listened with headphones and responded on the computer keyboard using a 4-level confidence scale (*sure same, same, different, sure different*). This is a change in comparison to the previous experiments in which we used Cakewalk to present a limited number of randomized lists to different groups of listeners.

RESULTS AND DISCUSSION

Area scores (see Table 1 and Figure 10A) were subjected to a 2 Delay X 2 Item Comparisons X 2 Music training ANOVA in which training was the only between-groups variable. The interaction of Delay X Item Comparison, $F(1, 29) = 41.08$, $p < .001$, $MSE = 0.002$, indicated an improvement in memory for comparisons between target (test) items and similar lures, $t(30) = 3.39$, $p < .05$, but not for comparisons between target (test) items and

different-contour lures. The three-way interaction of those variables with music training was significant, $F(1, 29) = 6.28$, $p < .02$, $MSE = 0.002$, but did not alter the qualitative pattern shown in Figure 10A, and did not bear on the issues under focus here. This three-way interaction showed the same pattern as the two-way interaction, but with a nonsignificant increase with delay in discrimination between target (test) items and different lures for untrained participants, and a corresponding nonsignificant decrease for moderately trained participants, coupled with slightly better overall performance by the moderately trained. There were no other significant effects.

Proportions of false alarms to similar lures declined with delay, $t(30) = 1.84$, $p < .05$ (one-tailed), as did false alarms to different lures, $t(30) = 2.97$, $p < .05$ (two-tailed; see Table 1 and Figure 10B).

As in Experiment 1, discrimination between target (test) items and similar lures improved with delay. Critically for this experiment, discrimination between copies of targets and different lures did not decline with delay, even when the *only* difference between different lures and target (test) items was their melodic contour. Therefore we can reject the alternative hypothesis that the observed improvement in discrimination between target (test) items and similar lures is due to a decline over time in the salience of contour as a retrieval cue, since the hypothesized decline in salience did not lead to poorer performance in discriminating changes in

contour. It is true, however, that discrimination between target test items and different lures was overall not better than discrimination between target test items and similar lures, as the manipulation reducing the difference between targets and different lures to just differences in contour, reduced the number of feature mismatches available.

Experiment 8

In Experiment 8, we tested the specific prediction of our feature binding hypothesis that discrimination between target (test) items and similar lures will improve with delay even when similar lures and targets have all their features in common except for the pitch level at which the contour is attached to the scale. If the improvement with delay in the rejection of similar lures is due to the binding of the contour to the appropriate degree of the tonal scale, then recognition memory performance should still improve when the *only* cue to the difference between similar lures and targets lies in the pitch level at which the melody is presented. In Experiments 1-3 and those of Dowling et al. (2002), there were additional uncontrolled cues, other than simply the pitch level of the melody on the scale, that differentiated targets from similar lures. See, for example, the contrast between brackets 1 and 3 in Figure 3. Note the differences in the bass lines, which involve a change in the harmony. The contrast of brackets 2 and 7 illustrates this point also (though the phrase in bracket 7 is not in the correct temporal delay position to serve as a similar lure with respect to the phrase in bracket 2). There again the bass lines and the harmony are different. All of these additional differences arose due to our initial aim, motivated by ecological validity, of presenting the music just as written by the composer. Therefore, in Experiment 8 we created similar lures that were exact copies of targets in every respect except for the pitch level of the melodic shape. For example, we could take the target at bracket 1 (Figure 3) and copy it to make a test item, but with the melody shifted as in bracket 3, with the accompaniment adjusted to preserve the harmony and texture of the original. This manipulation provides a direct test of our binding hypothesis, since if this difference in pitch level along the scale is the critical feature that listeners are using to discriminate similar lures from targets at the long delay, they should still be able to do that when that is the *only* cue available. In order to concentrate on the assessment of discrimination between target (test) items and similar lures we did not include different lures in this experiment.

METHOD

Participants. Forty participants served in Experiment 8; 18 were moderately trained and 22 untrained.

Stimuli. Stimuli consisted of a new sample of 64 minuets, and were constructed as in Experiment 7, except that test comparisons were all either target (test) items or similar lures, and similar lures were copies of targets in all respects except that the contour of the melody was shifted along the scale to a different pitch level, as close as possible to the pitch level of the target while still preserving the original harmony. We used Cakewalk software to copy the target onto the similar lure test item, and then altered the pitches of the melody to change its pitch level but left the contour intact and in conformity with the harmony of the target.

There were 64 trials, recorded as .wav files and presented as in Experiment 7 by means of a Matlab program that produced a different random order of trials for each participant, who listened using headphones and responded on the computer.

RESULTS & DISCUSSION

Area scores (see Table 1 and Figure 11A) were subjected to a 2 Delay X 2 Music training ANOVA in which music training was the between-participants variable. Moderately trained participants performed better than untrained (.70 vs. .63), $F(1, 38) = 6.75, p < .02, MSE = 0.017$. The interaction of Delay X Training, $F(1, 38) = 11.52, p < .01, MSE = 0.007$, showed that discrimination between target (test) items and similar lures improved for moderately trained participants, $t(17) = 3.02, p < .05$, but not for the untrained, whose decline is not significant. There were no other significant effects.

Proportions of false alarms to similar lures declined with delay for moderately trained participants, $t(17) = 2.91, p < .05$, but not for untrained (see Table 1 and Figure 11B).

The results for moderately trained listeners showed improvement with delay in agreement with our binding hypothesis, whereas those for untrained listeners did not show improvement. The result for moderately trained listeners is in agreement with previous evidence showing that such listeners encode the pitches of novel melodies in terms of tonal scale degrees in other melody recognition tasks (Dowling, 1986), showing that they were encoding the position of the melodic contour on the scale for each novel melody they encountered.

It should be noted that performance for the untrained listeners did not decline with delay. This suggests that they were using the scale information and feature binding to some extent, since in the absence of some further

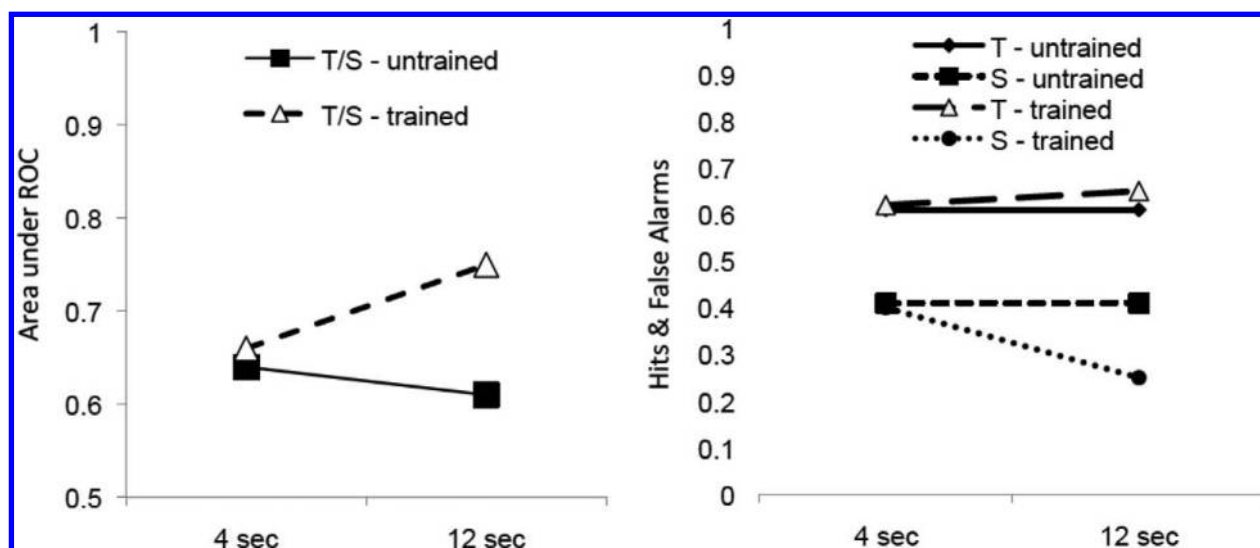


FIGURE 11. Results of Experiment 8, in which similar lures were copies of targets except that the melody was shifted up or down in pitch in relation to the scale. Here there was an interaction with experience. (Left) Area under the ROC showing unbiased proportion correct discrimination of target test items from similar lures (T/S) at short and long delays, for listeners untrained in music (untrained) and with moderate training (trained). (Right) Proportions of hits to target test items (T) and false alarms to similar lures (S) at short and long delays, for listeners untrained in music (untrained) and with moderate training (trained).

encoding we would not expect their performance to improve, as in Experiments 4-6. That this was the case is suggested by the fact that in Experiment 1 and Dowling et al. (2002) their performance improvement with delay was indistinguishable from that of moderately trained listeners. This also suggests that in those other studies in which the available stimulus features were not so tightly controlled, the untrained listeners were able to use the continued encoding of those other features in distinguishing between target (test) items and similar lures at the delayed test, where they generally performed better than at the shorter delay.

GENERAL DISCUSSION

There is considerable evidence that melodic contour is an important feature in recognition memory, and can be recognized independently of the precise pitch content of the melody (Dowling, 1978, 2001; Idson & Massaro, 1978; Jones et al., 1987; Kim & Levitin, 2002). In Dowling's (1978) framework, the precise pitch content of the melody is provided by a second feature of the melody: its tonal scale. The scale is a hierarchically organized set of pitches (Krumhansl, 1990; Lerdahl, 2001) that provides a framework for the listener's perception of distances along the pitch dimension, as Helmholtz (1877/1954) suggested. The scale is specified by the key of the melody; for the minuet in Figure 3 the key is G major. Dowling (1978) theorized that the pitch

pattern of a scale functions as a perceptual invariant common to numerous melodies, and that an actual melody consists of the melodic contour (including the relative sizes of pitch intervals coded as one or two diatonic steps and leaps) applied to a specific scale at the appropriate relative pitch level. For example, the song Frère Jacques can be sung using a major or minor scale, using the same contour but changing the scale. The tune will be recognizable, but the nuances of the tonal functions will be changed, thus changing the "dynamic shape" in Jones's sense (Jones et al., 1987).

In the present study, we investigated a memory improvement effect that has been observed for musical material occurring over delays filled with interfering material, namely, the continuation of the piece. We proposed an explanation in terms of feature binding, specifically of the melodic contour to a particular pitch level in the scale. When tested at short delays, listeners confuse target (test) items and similar lures, which share the individual features of melodic contour and musical key (scale), because they have not yet bound the contour to the scale at the appropriate pitch level and so are left without a means of differentiating the target (test) and the similar lure. At the longer delay, listeners accomplish this binding, and are better able to reject similar lures. We believe this binding takes so long to accomplish because the listeners are trying to follow the ongoing music and encode it while their memory systems are

still completing the encoding of earlier stimuli. Experiment 3 of Dowling et al. (2002—see Table 1) shows that if listeners are freed from having to follow the continuing music (that is, when the delay is filled with silence), the requisite binding is accomplished very rapidly. But even though slowed by the processing of the continuing music, binding continues while listeners process new interfering information, and so discrimination between target (test) items and similar lures improves at the longer delays. Note that one consequence of binding the target melody to the scale is that the differences of tonal functions of the various scale degrees is made more salient, which makes feature mismatches based on them more obvious when the listener is presented with a similar lure after the longer delay.

In the present experiments, we found that for this binding process to occur, the similarity of the filler material with the preceding, to-be-memorized material, as well as its continuity and coherence, is important. If similarity, continuity, and coherence are disrupted so that listeners hear the interpolated material as coming from a different source in the sense of Bregman's (1990) auditory scene analysis, the memory-improving binding processes no longer take place, and performance is either stable or decreases over time.

It may seem counterintuitive that in many of the cases where performance improves in terms of target test item vs. similar lure discrimination, hit rates to the target test items do not increase, and often decline. Why doesn't the hypothesized improved representation of the target produced by binding lead to higher hit rates as well as reduced false alarms? There are at least two possible factors operating here: the effect of the bound representation on feature matches and mismatches, and shifts in response bias. Constant hit rates across the delay may mean that the bound representation is especially useful in generating mismatches with similar lures, but doesn't add much benefit to the matches already produced with target test items that were based on the individual features. Hit rates can also decline if a shift toward greater caution in producing positive responses occurs at the longer delay interval. In that case all three response indices will show a decline, as in the results of Experiments 1, 2, and 7 (Table 1). In those cases what is important for discrimination is that false alarms decline more rapidly than hit rates. We should note, also, that given the methods we are using, it is impossible to distinguish a shift in response bias from an overall decline in perceived familiarity with increasing delay. It may be that in some cases the familiarity of target test items decreases over time, but just more slowly than the familiarity of lures.

The present phenomenon with music memory represents a case of binding processes involving complex features and nonverbal material. It provides a contribution to binding research, which has hitherto mainly focused on the visual modality (objects, scenes, words) or simple feature combinations of sounds (for example, pitch and timbre, as demonstrated by Hall, Pastore, Acker, & Huang, 2000). Binding in perception and memory has received considerable attention, as have its underlying neural correlates (see the volume by Zimmer, Mecklinger, & Lindenberger, 2006, for an overview). Different levels of binding can be differentiated: for example, low-level perceptual feature binding within items *versus* the binding of more complex constellations of features that lead to between-item binding. Zimmer et al. (2006, Chapter 1; see also Robertson, 2003) point out that such binding is not homogenous, but that bindings of locations and spatial context, or of temporal context, appear to differ with different types of material. They thus emphasize the importance for future research to address different modalities, tasks, and methods. The present study with the complex auditory patterns of music extends our understanding of binding processes to a new domain.

The processes reported here concern binding in the auditory modality; they concern nonverbal material and require not only bottom-up processes, but top-down processes related to listeners' knowledge of tonal structure, which facilitates the perception of musical structures and tonal functions. Bharucha (1987) describes how the listener's schematic knowledge of the tonal system generates expectations, which influence perception and facilitate encoding of patterns that follow the expected structure, including the binding of features into integrated perceptual representations. In the present experiments, filler material schematically related to the target helps to keep relevant sets of features active in working memory, facilitating the ongoing process. For example, the expected pitches of the tonal scale in the current key in a piece are processed more quickly and accurately than unexpected, nonscale pitches, and a similar facilitation occurs for notes at expected times (Dowling, 2001). The activation of schematic knowledge of the tonal scale is especially important in the binding of the melodic contour to the scale, which, as described above in connection with Figure 3, is particularly relevant to discrimination between target (test) items and similar lures. In this view, when target (exposition) and filler are continuously linked, they both activate schematic knowledge that aids in the binding of target features, with activation becoming stronger as the piece progresses. This activation aids in the

development of a memory representation that supports improved recognition performance. In contrast, filler material that departs too strongly from the target fails to continue to activate the relevant schematic knowledge, thus hindering the ongoing binding process.

In most music cognition research, memory is usually tested in same-different paradigms with musical excerpts being presented in isolation and with the to-be-compared items being separated by a silent interval. These silent retention periods thus allow for rapid binding of features. Working with continuous listening paradigms (that is, where the to-be-remembered item is not indicated explicitly and the retention interval is filled with the continuation of the music) thus brings to light the underlying binding processes, slowed down, but still functioning. The continuous listening situation not only brings us closer to real-world listening, but gives us the possibility of studying how complex relational binding can occur in spite of on-going processing of similar material. And the present results show that this binding can proceed only when the material being processed in parallel is perceived as part of the original to-be-processed object, the original sound source.

This feature binding that continues while the listener actively attends to the continuation of the music, and that leads to improvement in discrimination between target (test) items and similar lures, is surprising in the light of results repeatedly found with memory for prose stories (namely, that discrimination between target test items and similar lures declines over time (Gernsbacher, 1990, see Figure 1; Sachs, 1967). It may be that listeners have considerable practice in extracting the gist of a story, and in discarding surface details that are irrelevant for overall meaning. We have found some evidence that poetry, in which details of rhythm and rhyme are more important than in prose, tends to resemble music more than it does prose in regard to short-term retention of surface details, as indicated by discrimination between target (test) items and similar lures that are paraphrases of the target (Tillmann & Dowling, 2007). Surface details may be more likely to be retained in poetry than in prose because rhythm and rhyme can be useful as mnemonic aids in remembering poetry (Rubin, 1995), and so in poetry, verbal content becomes bound to patterns of rhythm and rhyme.

If it is true that the binding of a melodic contour to a tonal scale is an essential part of how we encode novel tonal music into memory, then it is surprising that our manipulations of tonality did not have a greater effect on performance. The upward shift of one semitone of the filler material in Experiment 3 perhaps lacked

impact because such shifts are common in popular music. But the rather large difference in key distance of the interpolated two measures in the minuets in Experiment 5 also had no measurable effect. It may be that the effect of a change in the thematic material in the interpolated minuet simply swamped the effect of key distance. Clearly further research is needed to separate the effects of tonality shifts from other sources of disruption.

The present results also have implications for the role of attention in binding processes. The binding of features into a coherent representation of a musical target phrase is apparently taking place while the listener is focusing attention on something else, namely subsequent phrases in the continuing music. Nevertheless, it is necessary that the listeners attend to the continuation of the same music, and not just any continuing music. In this regard, the effects of attention on binding are similar to those of attention on the formation of auditory streams. When a tone sequence skips rapidly among separate pitch regions, the pattern forms separate perceptual streams even without attention, a process that takes time (Sussman, Horvath, Winkler, & Orr, 2007). However, when attention is directed elsewhere, it can still matter where attention is directed; in the case where there are three possible streams, focusing attention on one of them inhibits the formation of separate streams for the other two (Sussman, Bregman, Wang, & Kahn, 2005). This shows that the control of encoding processes by attention that we have observed here is not an isolated case.

Up to now our studies have been entirely on the behavioral level. However, these phenomena and the paradigm for their investigation opens up the possibility of new ways of studying the underlying neurophysiology of feature binding processes in audition. As Robertson (2003) indicates, there is no single neural mechanism for binding that operates across senses and across levels of feature complexity. Thus, complex feature binding with musical patterns offers the possibility of a different perspective than that provided by other materials. We expect that hippocampal structures may be involved in complex feature binding of musical material, as they were with verbal materials (Kroll, Knight, Metcalf, Wolf, & Tulving, 1996), and also as suggested by functional imaging research showing increased activation of the hippocampus for feature binding of objects and locations (Mitchell, Johnson, Raye, Mather, & D'Esposito, 2000). It may be that hippocampal lesions would slow binding in the present paradigm even further, but an intriguing question is whether with music it could be that right hippocampal lesions were more important

than the left ones that slowed verbal processing (see, for example, Wanatabe, Yagishita, & Kikyo, 2008). The hippocampal patients might also show deficits in the silent-interval condition (Dowling et al., 2002, Experiment 3), specifically lower overall performance levels at short delays. There is also the possibility to consider parietal cortex involvement, since that region appears to be implicated in the binding of features to spatial locations (Robertson, 2003).

The binding hypothesis for our music material, taken together with binding research on other types of materials, leads to strong, testable predictions not only for the neural correlates just discussed, but also for the influence of aging. Age-related binding deficits have been previously demonstrated for words and objects and locations (see for example Lyle, Bloise, & Johnson, 2006, for a review), and Cabeza (2006) suggests a binding deficit as one of the principal sources of memory impairment in aging. At present there is little evidence concerning binding deficits in aging with musical materials. We expect a binding deficit would cause greater problems with discrimination between target (test) items and similar lures (which requires binding) than with discrimination between target (test) items and different lures (which simply requires feature matching). Clearly the next step in exploring this aspect of aging and cognition is to explore the present paradigm with older listeners.

In summary, the results suggest that the process of encoding music in memory is very different from the process of encoding prose. With prose we are well practiced in extracting the gist of the meaning, and in discarding surface details of sound patterns and temporal order as soon as they have served their purpose in facilitating that encoding. With music, the structured patterns

of sound are important, and in fact constitute the essence of the music. During continuous listening, the brain needs appreciable time, on the order of seconds, to encode these sound structures into coherent representations that take account of the multiple dimensions in which the music is constructed—dimensions of pitch and time, and of timbre and loudness and textural complexity—and notably to connect the pitch dimension to the tonal framework (that is, the scale). One reason this encoding process takes so much time is that the ongoing music itself makes demands on the listener's attention and itself requires attentional resources in turn. And the present data show that the perceived continuity of that ongoing music with what came before is essential to the continuation of the encoding of earlier material and the binding of individual stimulus features into coherent memory representations.

Author Note

We thank James Bartlett, Emmanuel Bigand, and Hervé Abdi for helpful suggestions in the design and interpretation of the experiments, and Brian Hurley, Rachel Kowert, Shannon Layman, Ramiro Lopez, Rachel Parra, Natalie Peterson, Vannak Sung, Sarah Vahedipour, and Staron Walker for assistance in stimulus generation, running the experiments, and data analysis. This research was supported in part by a grant from the ANR "Music and Memory."

Correspondence concerning this article should be addressed to W. Jay Dowling, School of Behavioral & Brain Sciences, University of Texas at Dallas, 800 West Campbell Road, Richardson, TX 75080-3021. E-mail: jdownling@utdallas.edu

References

- BEETHOVEN, L. VAN (1967). *Klaviersonaten* (Vol. 1). Munich: G. Henle.
- BEETHOVEN, L. VAN (1987). *Bagatelles, rondos, and other shorter works for piano*. New York: Dover.
- BEETHOVEN, L. VAN (1990). *Tänze für Klavier*. Munich: G. Henle.
- BHARUCHA, J. J. (1987). Music cognition and perceptual facilitation: A connectionist framework. *Music Perception*, 5, 1-30.
- BREGMAN, A. S. (1990). *Auditory scene analysis*. Cambridge, MA: MIT Press.
- CABEZA, R. (2006). Prefrontal and medial temporal lobe contributions to relational memory in young and older adults. In H. D. Zimmer, A. Mecklinger, & U. Lindenberger (Eds.), *Handbook of binding and memory: Perspectives from cognitive neuroscience* (pp. 595-626). Oxford, UK: Oxford University Press.
- DEWITT, L. A., & CROWDER, R. G. (1986). Recognition of novel melodies after brief delays. *Music Perception*, 3, 259-274.
- DOWLING, W. J. (1978). Scale and contour: Two components of a theory of memory for melodies. *Psychological Review*, 85, 341-354.
- DOWLING, W. J. (1986). Context effects on melody recognition: Scale-step versus interval representations. *Music Perception*, 3, 281-296.
- DOWLING, W. J. (2001). Music perception. In E. B. Goldstein (Ed.), *Handbook of perception* (pp. 469-498). Oxford, UK: Blackwell.
- DOWLING, W. J. (2010). Music perception. In C. Plack (Ed.), *Oxford handbook of auditory science: Hearing* (pp. 231-248). Oxford, UK: Oxford University Press.

- 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.-Y., & ANDREWS, M. W. (1995). The time course of recognition of novel melodies. *Perception and Psychophysics*, 57, 197-210.
- DOWLING, W. J., TILLMANN, B., & AYERS, D. F. (2002). Memory and the experience of hearing music. *Music Perception*, 19, 249-276.
- GABRIELSSON, A. (1999). The performance of music. In D. Deutsch (Ed.), *The psychology of music* (pp. 501-602). San Diego, CA: Academic Press.
- GERNSBACHER, M. A. (1990). *Language comprehension as structure building*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- GERNSBACHER, M. A. (1997). Two decades of structure building. *Discourse Processes*, 23, 265-304.
- HALL, M. D., PASTORE, R. E., ACKER, B. E., & HUANG, W. (2000). Evidence for auditory feature integration with spatially distributed items. *Perception and Psychophysics*, 62, 1243-1257.
- HAYDN, J. (1984). *Piano sonatas* (2 volumes). New York: Dover.
- HAYDN, J. (1989). *Tänze für Klavier [Dances for the keyboard]* (Hob. IX: No. 3, 8, 11, 12). Vienna: Schott.
- HELMHOLTZ, H. L. F. VON (1954). *On the sensations of tone*. (A. J. Ellis, Trans.) New York: Dover. (Original work published 1877)
- IDSON, W. L., & MASSARO, D. W. (1978). A bidimensional model of pitch in the recognition of melodies. *Perception and Psychophysics*, 24, 551-565.
- JONES, M. R., SUMMERELL, L., & MARSHBURN, E. (1987). Recognizing melodies: A dynamic interpretation. *Quarterly Journal of Experimental Psychology*, 39A, 89-121.
- KIM, J.-K., & LEVITIN, D. J. (2002). Configural processing in melody recognition. *Canadian Acoustics*, 30, 156-157.
- KROLL, N. E. A., KNIGHT, R. T., METCALFE, J., WOLF, E. S., & TULVING, E. (1996). Cohesion failure as a source of memory illusions. *Journal of Memory and Language*, 35, 176-196.
- KRUMHANSL, C. L. (1990). *Cognitive foundations of musical pitch*. New York: Oxford University Press.
- KRUMHANSL, C. L., & SHEPARD, R. N. (1979). Quantification of the hierarchy of tonal functions within a diatonic context. *Journal of Experimental Psychology: Human Perception and Performance*, 5, 579-594.
- LAKOS, A. (ED.) (1994). *Alte Tänze [Old dances]*. Budapest: Kunemann.
- LERDAHL, F. (2001). *Tonal pitch space*. New York: Oxford University Press.
- LYLE, K. B., BLOISE, S. M., & JOHNSON, M. K. (2006). Age-related binding deficits and the content of false memories. *Psychology and Aging*, 21, 86-95.
- MITCHELL, K. J., JOHNSON, M. K., RAYE, C. L., MATHER, M., & D'ESPOSITO, M. (2000). Aging and reflective processes in working memory: Binding and test load deficits. *Psychology and Aging*, 15, 527-541.
- MOZART, W. A. (1956). *Sonatas and fantasies for the piano*. Bryn Mawr, PA: Theodore Presser.
- MOZART, W. A. (1992). *Sonatas and variations for violin and piano* (2 volumes). New York: Dover.
- ROBERTSON, L. C. (2003). Binding, spatial attention and perceptual awareness. *Nature Reviews Neuroscience*, 4, 93-102.
- RUBIN, D. C. (1995). *Memory in oral traditions: The cognitive psychology of epic, ballad, and counting-out rhymes*. New York: Oxford University Press.
- SACHS, J. (1967). Recognition memory for syntactic and semantic aspects of connected discourse. *Perception and Psychophysics*, 2, 437-442.
- SCHUBERT, F. (1989). *Dances for solo piano*. New York: Dover.
- SUSSMAN, E. S., BREGMAN, A. S., WANG, W. J., & KAHN, F. J. (2005). Attentional modulation of electrophysiological activity in auditory cortex for unattended sounds in multistream auditory environments. *Cognitive, Affective and Behavioral Neuroscience*, 5, 93-110.
- SUSSMAN, E. S., HORVATH, J., WINKLER, I., & ORR, M. (2007). The role of attention in the formation of auditory streams. *Perception and Psychophysics*, 69, 136-152.
- SWETS, J. A. (1973). The relative operating characteristic in psychology. *Science*, 182, 990-1000.
- TILLMANN, B., & BIGAND, E. (1998). Influence of global structure on musical target detection and recognition. *International Journal of Psychology*, 33, 107-122.
- TILLMANN, B., BIGAND, E., & MADURELL, F. (1998). Local versus global processing of harmonic cadences in the solution of musical puzzles. *Psychological Research/Psychologische Forschung*, 61, 157-174.
- TILLMANN, B., & DOWLING, W. J. (2007). Memory decreases for prose but not for poetry. *Memory and Cognition*, 35, 628-639.
- TILLMANN, B., DOWLING, W. J., MOLIN, P., LALITTE, P., SCHULZE, K., POULIN-CHARRONNAT, B., SCHOEN, D., & BIGAND, E. (2013). Influence of expressive versus mechanical musical performance on short-term memory of musical excerpts. *Music Perception*, 30, 419-425.
- 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' , Az , and A' . *Perception and Psychophysics*, 68, 643-654.
- WANATABE, T., YAGISHITA, S., & KIKYO, H. (2008). Memory of music: Roles of right hippocampus and left inferior gyrus. *Neuroimage*, 39, 483-491.
- ZIMMER, H., MECKLINGER, A., & LINDENBERGER, U. (2006). *Handbook of binding and memory: Perspectives from cognitive neuroscience*. New York: Oxford University Press.