

Memory Improvement With Wide-Awake Listeners and With Nonclassical Guitar Music

W. Jay Dowling and Heidi Magner
University of Texas at Dallas

Barbara Tillmann
Lyon Neuroscience Research Center

Our previous research demonstrated memory improvement for phrases of classical minuets across delays of 4–15 s during which listeners heard the continuation of the piece (*Music Perception*, 2002, 2014). This improvement was especially strong for memory of fine details of the music, assessed in terms of discrimination between targets and similar lures (phrases that retained the melodic and rhythmic contours of targets, but shifted them in pitch). We attributed this improvement to continued encoding of the target while the attended music continues. Here we replicate that improvement effect and extend it to a style other than classical music, using Ottmar Liebert's guitar music. On each trial, listeners heard the beginning of a piece, with a target phrase near its start. The piece continued for 3, 6, 12, or 24 s, after which there was a "beep" and a test item: a copy of the target (T), a similar lure (S), or a different lure (D). We assessed memory in terms of both T/S and T/D discrimination. T/S discrimination improved with increasing delay up to 12 s, but only when listeners were tested near the peak of their circadian rhythm (vs. far from the peak). These results suggest that the continued encoding of targets requires attentional resources, and further shows that the previously observed memory improvement effect is not restricted to classical music.

Keywords: memory, music, binding, encoding, circadian rhythms

Recent results with memory for real pieces of music have shown improvement in recognition performance over delays of 3–15 s after the initial presentation of a musical phrase (Dowling & Tillmann, 2014; Dowling, Tillmann, & Ayers, 2001; Tillmann et al., 2013). In a typical study, listeners heard the beginning of a classical minuet. Toward the start of the minuet, there was a target phrase that would be tested later. The minuet continued just as the composer had written it, without repeating or imitating the target phrase. After a shorter or longer delay filled with the continuation of the minuet, the listener heard a high-pitched "beep" that signaled the onset of the test. The test phrase that followed was the same as the target phrase (T), a similar lure (S) sharing the melodic-rhythmic contour of the target, but differing from it in detail (typically a shift in pitch of the melodic line), or a different lure (D) having a different melodic-rhythmic contour. Listeners had to decide whether the test phrase was exactly the same as an earlier phrase they had heard in that piece, or different. Discrimination of targets from similar lures (T/S discrimination) improved with delay between 3 s and 15 s. This improvement was principally due to a decrease in false alarms to similar lures while hit rates remained constant over time. That is, with increasing delay, listeners improved their discrimination of targets and similar lures.

The first three experiments shown in Table 1 provide some typical results from those experiments.

We attributed this pattern of responses to the continued encoding of the musical phrases while listeners continued to follow the piece, binding together separate features encoded individually when the listeners had first heard the target phrase (Dowling & Tillmann, 2014; Tillmann & Dowling, 2007; Tillmann et al., 2013). For example, the melodic contour and the musical scale in a particular key would be encoded separately as individual features early in the memorization process.¹ When a decision concerning an S lure test item is made at the shorter delay, the memory system would rely on the considerable overlap of individual features between T and S, and accept the similar lure as the same as the

¹ We are focusing on the binding of the features of contour and scale, and not, for example, on contour and melodic interval pattern, because there is evidence for the psychological reality of the scale, and relatively little for the interval pattern (Dowling, 1991; Dowling, Kwak, & Andrews, 1995). Krumhansl (1990) has amassed considerable evidence for the stability of listeners' judgments regarding the tonal hierarchy (i.e., the scale), and for the relevance of the tonal hierarchy to listeners' perception of music. Deviations from the current scale pattern in a melodic stream of notes are easily noticed (Janata, Birk, Tillmann, & Bharucha, 2003). Novel melodies that conform to overlearned scale patterns are easier to remember than those that do not (Dowling & Fujitani, 1971; Francès, 1958/1988; Schulze, Dowling, & Tillmann, 2012). Dynamic tendencies of pitches in the tonal hierarchy (such as the tendency of the leading tone to the tonic) are properties of scale steps, and not of intervals. That is, inverted intervals involving the same pitch classes, such as thirds and sixths, are heard as quite similar (Balzano & Liesch, 1982). And the interval pattern of a familiar melody can be destroyed by scrambling its pitch classes into several octaves or by temporally interleaving distractor notes, and the melody is still recognizable in cued recognition (Dowling, 1984; Dowling, Lung, & Herrbold, 1987).

This article was published Online First February 11, 2016.

W. Jay Dowling and Heidi Magner, School of Behavioral & Brain Sciences, University of Texas at Dallas; Barbara Tillmann, Centre National de la Recherche Scientifique, Lyon Neuroscience Research Center.

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. E-mail: jdowling@utdallas.edu

Table 1
Area Under the MOC for T/S and T/D Performance in Previous Experiments With Various Delay Times (in seconds)

Experiments	Delay (s)					
	2.5	4	5	9.5	12	15
Dowling et al. (2002) Expt 1						
T/S			.61			.78
T/D			.78			.81
Dowling & Tillmann (2014) Expt 1						
T/S		.69			.76	
T/D		.80			.80	
Tillmann et al. (2013) – expressive						
T/S	.59			.76		
T/D	.88			.76		
Dowling et al. (2002) Expt 3						
T/S			.87			.86
T/D			.97			.95

target (leading to higher false-alarm rates). After a longer delay, during which additional encoding of the target phrase (and notably, the binding together of the features) continues, the contour and the scale would become bound together. This binding would produce a coherent memory representation of the phrase that could be stored as a unit in memory, retrieved when needed, and manipulated in working memory. Because of the binding of the individual features, differences between the S lure and the target would become apparent and the S lure would be rejected (leading to a decreased rate of false alarms). An example of the contour of a target phrase is shown in Figure 1. The contour of an S lure is labeled S. The pitches of this contour have been shifted up or down, while the contour has maintained the same shape. When

tested at the short delay, the individual features of contour and key are sufficient for the memory system to return a positive response in comparing the S lure with the target (i.e., a false alarm). However, after a longer delay, and thus after additional encoding in which the contour is bound to the scale at the appropriate pitch in the representation of the target, the S lure is easier to reject and a false alarm is more easily avoided. It is only when the contour and tonal scale are combined, with the contour being bound to the scale at a particular pitch level, that the particular musical structure underlying the phrase is clarified (Dowling, 1978). In summary, when memory is tested before the contour and scale features have been bound together, there is confusion between Ts and Ss, as they share a high proportion of individual features. Later, when the features have been bound, the differences between the target and similar lure phrases become apparent.

The memory representation in which the features are bound together plays a role similar to that of Treisman's "object file" in her Feature Integration Theory of visual encoding (Treisman, 2006; Treisman & Zhang, 2006). Such a memory representation can be stored and retrieved using working memory, can be compared with other similar representations of phrases, and can be manipulated in musical imagery. As in the Feature Integration Theory, we hypothesize that the initial encoding of features, such as the encoding of contour and scale, is largely automatic, requiring little or no attentional resources. There is some evidence for the automatic encoding of contour elements in the literature, notably measuring the mismatch negativity in auditory evoked responses (Koelsch, 2011; Koelsch & Siebel, 2005). However, it seems very likely that, as in vision, feature binding (here of contour and scale) requires attentional resources (Treisman, 2006). Indirect evidence for this has been provided by Experiment 3 of Dowling et al.

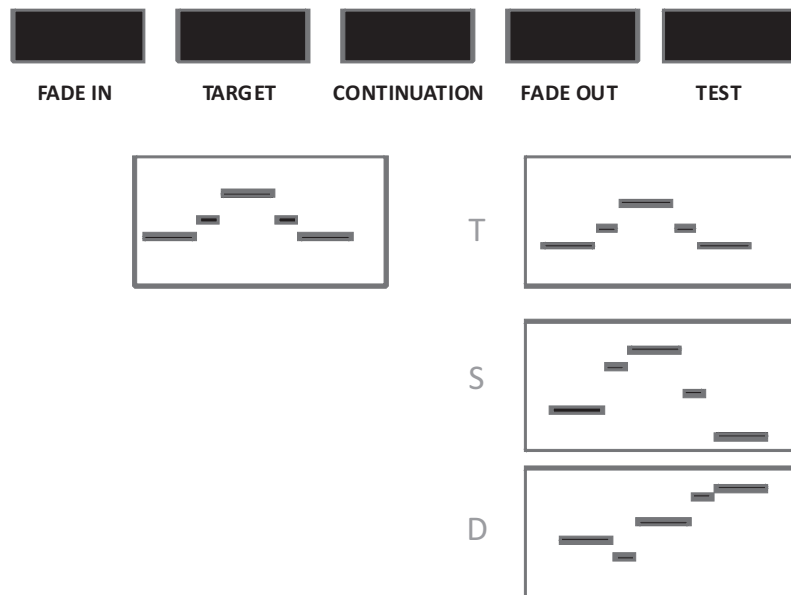


Figure 1. The sequence of events on a trial. The music fades in and presents one or two phrases leading directly to the target, which is followed by the continuation of the music for the time specified by the delay condition. Then the music fades out briefly, and the test phrase is presented. Sample melodic contours of a target and its test phrases are shown. The test could repeat the target (T), or have the same contour but with the notes shifted in pitch (S), or have a different contour (D).

(2001). In this experiment, we eliminated the continuation of the minuet in the delay interval between target and test, substituting a silence of short or long duration. This removed the additional processing of the continuing music (and the required attentional load) while performing the continued encoding of the target, and the results were equally good for T/S discrimination at the short delay and at the long delay (see the fourth experiment presented in Table 1). When listeners were free to devote a full complement of attentional resources to encoding the target, this was done very quickly, and the relevant features were bound together by the time of the short-delay test.

The present study provides a more direct test of the attentional resource requirements of continued encoding and feature binding. It is based on an extensive literature on the effects of circadian rhythms on attention and memory tasks. A reliable questionnaire developed by Horne and Östberg (1976)—the morningness/eveningness questionnaire (MEQ)—places people on a continuum between those whose circadian peak functioning is in the morning and those whose peak is in the afternoon and evening. (See Hasher, Goldstein, & May, 2005, for a review of the background literature and the reliability of the questionnaire.)

We tested listeners in the early morning or in the late afternoon and evening to investigate whether the synchrony between their circadian peaks and the time of testing would affect their performance. Based on the results from the work on the Feature Integration Theory (Treisman, 2006), we hypothesized that the continued encoding and feature binding, which require attentional resources, should be more pronounced for listeners tested near their circadian peaks than for those tested at off-peak times (what Hasher et al., 2005, call a *synchrony* effect). That is, we should see improved T/S discrimination with delay in listeners tested on peak, and less (or no) improvement with delay in listeners tested off peak.

In addition to using circadian synchrony to assess the role of attentional resources in feature binding, we also wanted to extend our investigation of these effects to (a) a different genre of music, (b) a larger set of delay intervals between target and test, and (c) different patterns of construction of test items and how these test items are integrated into the ongoing music. Previously, we had used classical minuets, with the first studies using tightly controlled, somewhat mechanically produced stimuli (Dowling & Tillmann, 2014; Dowling et al., 2001), and had then shown that the effect generalizes to expressive performances produced under natural conditions by an expert pianist (Tillmann et al., 2013).² (The similarity of the results can be seen in Table 1, in the first two experiments compared with the third experiment.) In the present study, we aimed to extend the effects to another musical style and used the popular guitar music of Ottmar Liebert, a guitarist born in Cologne and now living in Santa Fe, New Mexico, who characterizes his music as “nouveau flamenco” (www.ottmarliebert.com).

Furthermore, in the present study, we used a greater variety of delay intervals between target and test than we had typically used before (3, 6, 12, and 24 s). Whereas in our experiments using MIDI-generated stimuli (Dowling & Tillmann, 2014; Dowling et al., 2001), we had used delays of 5 and 15 s or of 4 and 12 s, when we turned to the naturally expressive stimuli, the short and long delays averaged 2.5 and 9.5 s (Tillmann et al., 2013), respectively, because of the natural sounding tempi

chosen by the pianist. The results with the expressive minuets showed an improvement in T/S discrimination from 2.5 to 9.5 s—.59 to .76 correct—that was very similar to that obtained with the more mechanical stimuli of Dowling et al. (2001) Experiment 1 with delays of 5 and 15 s (.61 to .78—see Table 1). And the improvement observed by Dowling and Tillmann (2014—see Table 1) with 4 and 12 s (for mechanical stimuli) was in the same range, though not as strong. In all three experiments, T/D discrimination remained roughly constant across delay and fell in the range 76% to 88%. In the present study, we wished to include a wider range of delays in the same experimental paradigm. In particular, we wanted to see whether circadian synchrony effects would persist beyond the 15-s delay that we had tested extensively in earlier studies, and also include a 6-s delay in the middle of the previously observed delay intervals. (In Experiment 2 of Dowling et al., 2001, we had tested also at a 30-s delay, and observed little change in performance between 15 and 30 s.) We decided on the logarithmically spaced series 3, 6, 12, 24 s, because this set covers delays used earlier (3–12 s) as well as a 6-s interval interpolated between those, and a longer interval (24 s) that is close to the longest interval (30 s) tested up to now. For technical reasons (explained below), we had different groups of listeners perform the experiment with delays of 3 and 12 s, and with delays of 6 and 24 s. We hypothesized that the 3–12-s results would qualitatively replicate those of previous studies (see the first three experiments in Table 1), at least for listeners tested near their circadian peaks, and that for the on-peak listeners, performance at 6 s would fall between performance at 3 and 12 s, despite being produced by a different group of listeners. We expected that the performance of off-peak listeners would be weaker than the performance of on-peak listeners with regard to the improved rejection of S lures over the first 12 s. But as to whether they might show a steady drop from 3 to 12 s, or a slight improvement, or an improvement for shorter delays (between 3 and 6 s), we had no specific predictions. We thought that T/S performance would remain strong at 24 s, as it did in the 30-s condition of Dowling et al. (2001), under the hypothesis that this is not influenced by the other differences among the experiments, involving differences in stimulus sets and styles of presentation.

In assigning the four delays to the two groups, we wanted each group’s test items to include a relatively short delay (3 or 6 s) and a relatively long delay (12 or 24 s), so that they would not simply come to expect only short delays or only long delays and adopt memorization strategies accordingly. Therefore, we assigned 3 and 12 s to one group and 6 and 24 s to the other group.

Finally, the use of the recordings of Ottmar Liebert pieces also led to another change in the implementation of the exper-

² As an appropriate control for making this comparison, the “mechanical” versions were matched in tempo to the natural ones. They were “mechanical” only in their rigid control of tempo and note onset times, and in using *t* test items that repeated targets exactly. They were relatively “natural” in their relative intensities and in the articulation (staccato-legato) of notes. They were thus not mechanical in the sense of MIDI versions that are simply literal transcriptions of the written score, with no nuances of accent, articulation, or textural balance.

imental paradigm. We needed to (a) introduce a brief silence after the material presented during the delay between the presentation of the target and the presentation of the test item, and (b) present a test that was not the natural continuation of the music as written (as it had been for the classical minuet material). Because of the formal properties of the classical minuets, we had been able to rely on patterns of repetition and imitation in the original pieces to provide for tests with Ts, Ss, and Ds at the appropriate delay intervals. With the guitar pieces taken from CDs recorded by Ottmar Liebert, it was not possible to use the appropriate test items in the temporal positions in which they appeared in Ottmar Liebert's performance, because (a) none of them appeared at appropriate times, and (b) counterbalancing (see Method) required that each of the pieces was to be presented with a test item occurring at each of the four delay intervals. This adaptation of the paradigm had two principal consequences for the relationship of target and test. First, as we simply spliced the appropriate test item into the ongoing piece at its appropriate delay interval after a short silence, the continuity of the filler material between the target and the test was broken. We had previously tested various disruptions of the filler material and found that obvious changes in the apparent source of the music (in the sense of Bregman's, 1990, auditory scene analysis, such as disruptions arising from changes in instrumental timbre or meter) disrupted continued encoding of targets (Dowling & Tillmann, 2014) and eliminated the improvement in T/S discrimination. Here, we were testing whether breaking the continuity in the melodic line and the harmony just at the start of the test item would be similarly disruptive. Based on the pattern of experiments in Dowling and Tillmann (2014) testing various types of disruption, we hypothesized that we would obtain the T/S improvement effect seen in the earlier studies, as the test items preserved the key, meter, style, and instrumentation of the preceding material, which was clearly from the same source as the beginning part of the excerpt; and furthermore, the break between filler and test was not abrupt, but involved a smooth fading out and in of the material. In addition, the typical S lure in our previous studies tended to consist of the same melodic contour shifted in pitch along the scale. In the present study, the available S items (i.e., phrases that were melodic imitations of targets) included a wider range of possible transformations of the target, in that variations in harmony often accompanied the shift of the pitch level of the melodic contour as shown in Figure 1. We investigated whether the same kinds of binding processes that were operating to produce T/S improvement with the more restricted set of target-test relationships of the minuets would operate as well under these varied conditions.

In summary, the present study extends the investigation of the phenomenon of the improvement of T/S discrimination with increasing delay to a new genre of music, very different from the classical minuets we have used up to this point. It also includes more delay intervals in the same paradigm than we had used before: four delays ranging from 3 to 24 s. And it introduces a more flexible manner of connecting the test item to the musical material filling the delay, making possible the introduction of any of the three test types (T, S, or D) after any delay in the music. In addition, to test the hypothesis of required attentional processes for the continued processing and binding

of the features, we tested participants either on or off their circadian peak.

Method

Participants

In all, 155 participants served in the experiment in return for partial course credit. We tested memory after four possible delays, organized in terms of two pairs of delays performed by two groups of participants, randomly assigned. Seventy-four participants performed the experiment with delays of 3 and 12 s, and 81 with delays of 6 and 24 s. Participants completed a brief questionnaire concerning their musical training, and approximately equal numbers of less trained and moderately trained participants served in each delay condition. The 78 moderately experienced participants had 2 or more years of individual music lessons or experience performing in an instrumental ensemble such as band or orchestra ($M = 5.3$ years); the 77 less experienced participants had less than 2 years of those activities ($M = .8$ years).

To test participants during their on- or off-peak times in relation to their circadian rhythms, we administered an MEQ designed by Horne and Östberg (1976) to determine whether they were "late night" or "early morning" people. Horne and Östberg developed an 81-point scale that could be coded into the categories "Definitely Evening" (16–30), "Moderately Evening" (31–41), "Neutral" (42–58), "Moderately Morning" (59–69), and "Definitely Morning" (70–86). We split our sample into evening and morning groups at the median scale value of 47 for our sample, thus including a number of listeners in the neutral category who merely tended in one direction or the other. (We repeated our statistical tests on the data leaving out those in Horne and Östberg's neutral category, obtaining qualitatively the same results.) We ran the experiment early in the morning (between 8 and 11 a.m.) or late in the afternoon (between 4 and 7 p.m.), and categorized the participants as to whether they were participating on-peak (92 participants) or off-peak (63 participants) relative to their circadian rhythms. To the extent possible, approximately equal numbers of participants in each circadian condition and each experience level served in each delay condition. As some self-selection was operating in the times for which participants signed up for the experiment, it would clearly be a good idea to replicate the experiment with random assignment of participants to participation times, which was not possible in the present case.

Stimuli

In constructing the stimuli, we drew on Ottmar Liebert's seven albums published during the period of 1989 to 2003 (Liebert, 1989, 1990, 1992, 1993, 1995, 2001, 2003). To be useful in our study, a track on an album had to have a distinctive melodic phrase that, once it occurred, was neither repeated nor imitated for at least 30 s following its introduction. Each target also required imitative phrases elsewhere in the piece that could serve as S lures, and a different phrase in the piece to serve as a D lure. We found 24 songs that met these criteria, each of which included a potential T, S, or D test item. We then assigned those songs randomly to trials and testing conditions (T, S, or D), assigning them differently in two counterbalanced lists.

The fact that we were able to find only 24 passages that met our criteria for use in this paradigm meant that we had to limit the number of trials that any group of participants would perform to 24, so that each participant would hear each novel stimulus only once. With two delays, six stimuli are required for each set of data points (i.e., a T, S, and D item at each delay). Twenty-four trials provide for four iterations of those six trial types. Therefore, we had two groups of participants perform the experiment, each at two delay intervals. One group performed the experiment with delays of 3 and 12 s, thus replicating delays close to the previous delays shown in Table 1. The other group performed the experiment at 6 and 24 s, filling in the point between 3 s and 12 s, and adding a longer delay than most of those used before.

T items involving repetition of the target were constructed by literally copying the target phrase into the test position following the appropriate delay. S items consisted of phrases of the same length as target items that imitated the target melody, preserving the overall melodic and rhythmic contour, that is, the pattern of ups and downs of pitch and the relative durations of the notes (see example S in Figure 1). These imitations preserved the melodic contour, meter, and tempo of the target; they generally differed in the pitch level of the melody and in harmony, and in slight changes in the accompaniment. D items had clearly different melodic and rhythmic contours from targets, as well as different chord patterns, but were drawn from the same piece and were of the same length as targets, and preserved the same meter and tempo. We intended T/D discrimination to assess memory for salient, easily encodable individual features of melodies, such as contour.

Once a target phrase was selected (with respect to the musical structure and the possibility of finding suitable test items), trials began approximately 2 or 5 s (randomly determined) before the onset of the target phrase with the music leading into it. We did that to avoid cuing the listener as to the identity of the target phrase, just as we did in our previous studies (Dowling & Tillmann, 2014; Dowling et al., 2001; Tillmann et al., 2013). We did not simply start each trial with the target, because in that case, the listener could simply concentrate on remembering the target and ignore the continuation of the music. For example, even with uncertainty as to which of the first two phrases would be tested, and even while showing T/S improvement with delay, listeners in Dowling et al. (2001) Experiment 1 performed much better in discriminating targets presented in measures 1 and 2 from Ss and Ds (.84) than targets presented in measures 3 and 4 (.66). And we were interested in testing memory for phrases in an actual piece of music to which the listener is actively attending, with the understanding that any phrase (at least in a certain region of the piece) might be tested later.

At the start of each trial, the onset of the music was faded in with a 50-ms linear onset ramp. Targets were 2–6-s-long. Following presentation of the target, the music continued for 3 to 24 s (depending on the delay condition), and then faded out for 20 ms and stopped for 1.5 s during which a brief signal (a “choink” sound on the guitar) indicated that the test item would follow. The test item was then presented, followed by a 6-s response interval.

There were 24 trials: 12 with a short delay (3 or 6 s) and 12 with a long delay (12 or 24 s), with four trials at each delay tested with T, S, and D test items. (The limitation in the number of available targets as well as the need to keep the experimental session relatively short limited the number of possible trials in the exper-

iment.) As explained above, the two groups of participants performed the experiment either with delays of 3 and 12 s, or with delays of 6 and 24 s. We constructed two counterbalanced lists for each group, in which a target that appeared with a given delay and test type (T, S, or D) on one list appeared with a different delay and test type on the other list. This was especially important, given the limited number of trials for each test-item type, to ensure that any memory effects observed across different delays were not be simply attributable to a fortuitous assignment of stimuli to conditions. Approximately the same proportion of participants in each group at each experience level and in each circadian condition performed the experiment with each version of the list.

Timing in the experiment was controlled by Cakewalk software running on a PC-type computer reproducing the music via a 24-bit soundboard at CD sampling rates.

Procedure

Listeners served in group sessions of two to six participants and heard the stimuli over loudspeakers at comfortable levels. The experimenter had each participant complete the musical experience questionnaire and the MEQ. Then the experimenter explained the task, providing examples of the three different kinds of test items, and emphasizing that participants were to respond positively only if the test item was exactly the same as a phrase they had heard earlier in that trial. They responded with pencil on an answer sheet using a six-point confidence-level scale with the categories: “Very Sure Same,” “Sure Same,” “Same,” “Different,” “Sure Different,” and “Very Sure Different.” Participants were seated around the room at tables along the walls facing outward, which reduced their tendency to consult each others’ response sheets. The experiment took approximately 12–18 min to complete, depending on the delay condition.

We calculated area under the memory-operating characteristic (MOC—an unbiased estimate of proportion correct where chance is .50—Swets, 1973) for each participant’s discrimination of Ts from Ss and of Ts from Ds, for each condition. Thus, each area score was based on eight measurements in the original data. A measure relatively free of the effects of response bias, such as area under the MOC, is especially desirable in a study where bias effects may arise from individual differences based on musical experience and on circadian rhythms. Across numerous experiments, we have found area scores less likely to be correlated with bias measures such as *c* than other performance measures such as *d'* (Verde, Macmillan, & Rotello, 2006). We also obtained the proportions of hits and false alarms indicated by “same” responses (levels 4, 5, and 6) to T, S, and D items.

Results

We analyzed the area scores (Figure 2) by means of a $2 \times 2 \times 2 \times 2 \times 2$ analysis of variance (ANOVA) with groups (3–12-s delay, 6–24-s delay), experience levels (low, moderate), and circadian levels (on-peak, off-peak) as between-participant factors, and delay (short, long) and test-item comparison (T/S, T/D) as within-participant factors. There was better performance for T/D discrimination (.80) than for T/S discrimination (.70), $F(1, 147) = 119.37$, $MSE = .039$, $total \eta^2 = .059$, $p < .01$. Note that the overall effect of the four delays is assessed via the interaction of

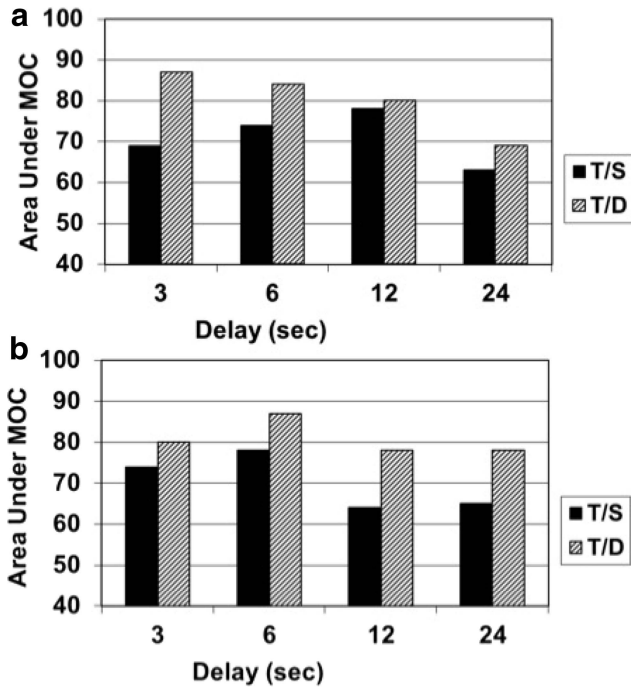


Figure 2. Area under the MOC for T/S and T/D discrimination for (a) two groups tested on their circadian peak at four delays: one group at 3- and 12-s delay, and the other group at 6- and 24-s delay, and (b) for similar groups tested off their circadian peak.

Group \times Delay. The effect of delay was significant, $F(1, 147) = 19.73$, $MSE = .039$, $total \eta^2 = .034$, $p < .01$, as was the Group \times Delay interaction, $F(1, 147) = 9.32$, $MSE = .039$, $total \eta^2 = .016$, $p < .01$. Overall performance rose slightly between 3- and 6-s delay, and declined thereafter, but the more important aspects of the results for delay involve interactions involving circadian synchrony.

The interactions of Group \times Circadian Level, $F(1, 147) = 5.09$, $MSE = .039$, $total \eta^2 = .012$, $p < .05$, and of Delay \times Circadian Level \times Test Item comparison, $F(1, 147) = 8.38$, $MSE = .039$, $total \eta^2 = .010$, $p < .01$, were significant; and their combined effects are shown in Figure 2. For participants performing the task on-peak, T/S discrimination rose between 3 and 12 s from .69 to .78, $t(37) = 3.89$, $p < .01$, as expected, and declined thereafter (Figure 2a). The data point for the T/S discrimination generated by the 6/24-s group tested after a 6-s delay (.74) fell in between the values for 3 and 12 s (.69 and .78) obtained with the other group, just as we expected on the basis of our previous results (Table 1). The difference between .74 and .78 approached significance with a one-tailed test, $t(90) = 1.49$, $p < .08$. For on-peak participants, T/D performance declined with increasing delay. For participants tested off-peak (Figure 2b), mean performance for T/S and T/D discrimination combined rose from 3 to 6 s (from .77 to .83, $t(61) = 5.43$, $p < .01$), but was lower thereafter (Figure 2b). No other effects were significant.

We analyzed hits and false alarms with a 2 Groups \times 2 Experience Levels \times 2 Circadian Levels \times 2 Delays \times 3 Test Items ANOVA, in which the last two variables involved within-participant comparisons (Figure 3). The effect of test item was

significant, $F(2, 294) = 433.51$, $MSE = .044$, $total \eta^2 = .39$, $p < .001$, indicating that participants distinguished the three kinds of items at above chance levels. Main effects of variables not interacting with test items simply indicate shifts of response bias, as with the effects of experience, $F(1, 147) = 3.92$, $MSE = .039$, $total \eta^2 = .002$, $p < .05$, and of delay, $F(1, 147) = 4.55$, $MSE = .039$, $total \eta^2 = .002$, $p < .05$, where there was a shift to more conservative response criteria with increased experience and delay. The interaction of Delay \times Test Item type was significant, $F(2, 294) = 10.13$, $MSE = .044$, $total \eta^2 = .010$, $p < .001$, indicating the change in recognition performance with delay noted above in the analysis of area scores. No other effects were significant, although the interaction of Group \times Delay \times Test Item approached significance, $F(2, 294) = 2.56$, $MSE = .044$, $total \eta^2 = .003$, $p < .08$.

Figure 3 shows that the improvement in T/S performance for on-peak participants arises largely from a steady decline from 3 to 12 s in false alarms to Ss, hit rates remaining roughly constant until the longest delay. This was not the case for off-peak performance, where S false alarms increased between 3 and 12 s. False alarms to Ds are similar for both on-peak and off-peak listeners, remaining relatively low with the highest rates occurring at the 12-s delay.

Discussion

The present results revealed an improvement in T/S discrimination, associated with a decline in S false alarms, during the first 15 s following the presentation of a novel musical phrase. This finding replicated with nonclassical guitar music our previous studies of recognition memory that used classical minuets (Dowl-

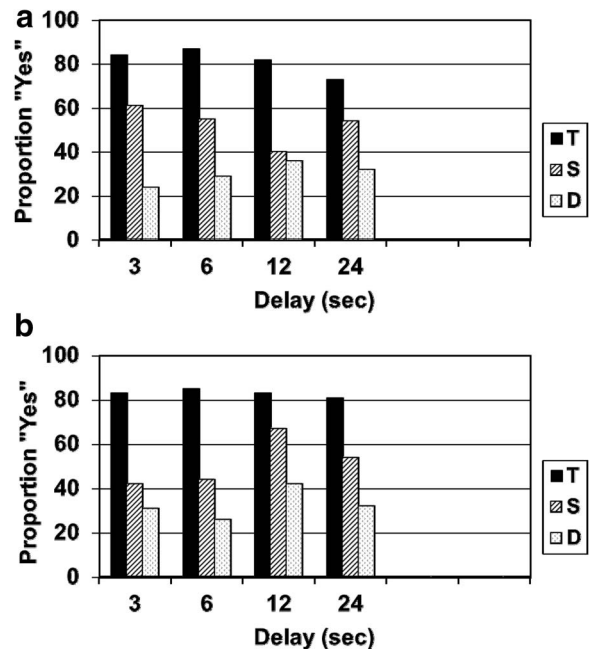


Figure 3. Proportions of hits (T) and false alarms (S, D) for (a) two groups tested on their circadian peak at four delays: one group at 3- and 12-s delay, and the other group at 6- and 24-s delay, and (b) for similar groups tested off their circadian peak.

ing & Tillmann, 2014; Dowling et al., 2001; Tillmann et al., 2013). It is thus clear that the improvement effect occurs across different genres of music that use very different instrumental textures, meters, and rhythms. Here we have found that this improvement effect was observed only for listeners tested on-peak in their circadian rhythm, suggesting that the additional encoding and feature binding, which we believe to be responsible for the memory improvement effect, require attentional resources whose availability depends on the circadian cycle (in which there are more attentional resources available “on peak”).

This finding supports our hypothesis, derived in parallel with Treisman’s (2006) Feature Integration Theory, that binding a melodic contour to a pitch level in the tonal scale requires attentional resources, just as binding does in other domains, such as binding a visual object to a location (Treisman & Zhang, 2006). Unlike our previous results, however, the improvement in T/S discrimination achieved at the 12-s delay did not persist at 24 s (Figure 2), as further discussed below.

Another hypothesis of the Feature Integration Theory (Treisman, 2006) is that, in contrast to the binding of the features, the encoding of individual features does not require attentional resources. Therefore, we would expect task performance that requires the matching of separate features to remain strong even for listeners tested off-peak, whereas task performance that requires the binding of the features would decline for these listeners tested off-peak. And this is exactly what we observed for participants tested off-peak (Figure 2b): at 12 and 24 s, T/D discrimination (which involves detecting a mismatch of individual features, such as melodic contour and rhythm) stays at around .78 (which is even better than the performance of the on-peak listeners), whereas T/S performance (which requires binding) descends into the low .60s.

The patterns of performance across the delay intervals were very different for on-peak and off-peak listeners. For off-peak listeners, T/S performance did not change much between 3 and 6 s, and declined thereafter; in particular, at the 12-s delay, it was 16 percentage points below the performance of on-peak listeners (Figure 2). It seems plausible that the off-peak listeners, lacking the attentional resources to engage in continued encoding, relied on strategies that emphasized success at the shorter delay intervals of the experimental session, whatever that interval was. Thus, both groups of off-peak listeners performed better at the shorter of their two delays (3 s or 6 s) than at the longer ones (12 s or 24 s). It would clearly be of value in future research to use other means of controlling the attentional resources available to listeners in this task, such as introducing a secondary task to occupy their attention. The dual-task approach would, we hope, provide converging evidence for the attentional demands of the convert memory processing (binding) we are studying.

One implication of the present results, given that our previous studies showed a strong T/S improvement effect even when we didn’t control for circadian synchrony (Table 1), is that participants in those earlier studies very likely tended to self-select for participation at times of day not too far from their circadian peak times. In general, the pattern of results in those earlier studies, in the conditions where we found the memory improvement effect during the first 15 s after hearing a novel piece of music, resembled that of the present on-peak participants much more closely than that of the off-peak participants. And in fact, as we did not administer the MEQ in a separate session and then randomly

assign participants to testing times, we found in the present study that participants were self-selecting for circadian synchrony (92 on-peak vs. 63 off-peak). This interpretation seems more plausible than attributing the interaction of the response pattern with circadian rhythms to the effects of the music of Ottmar Liebert. Future research might replicate our on/off-peak testing design with classical minuets and other material to confirm this hypothesis, and do it with random assignment of participants to times of participation.

These results, as well as those of our previous studies (Dowling & Tillmann, 2014; Dowling et al., 2001; Tillmann et al., 2013), converge with those of Edworthy (1985), who varied the length of to-be-remembered melodies from 3 to 15 notes, presented at a rate of 2 notes/s. There was a 5-s delay between the presentation of a novel melody and its test melody, and Edworthy measured response times for the detection of either changes of contour or changes in relative pitch of one of the notes. (All test items were transposed.) The detection of contour changes remained roughly constant or declined across the delay between the end of the target pattern and the altered note, a pattern similar to what we found with T/D discrimination, which involves detecting changes in contour. And detection of pitch changes improved with increased delay, just as we found with T/S discrimination.

As one of our eight S lures was an exact transposition of the target, the question arises, does the improvement over time in the rejection of S lures apply equally to S lures that are exact transpositions when shifted along the scale, as to S lures whose intervals change when they are shifted along the scale (as was the case in Dowling & Tillmann, 2014; Dowling et al., 2001, and Tillmann et al., 2013). That is, would we get the same effects using S lures in which none of the pitch intervals between notes is changed? Something close to this experimental test has already been done in previous work. Dowling (1986) presented targets that could be shifted along the scale between the tonic (first degree) and the dominant (fifth degree) without changes in melodic intervals. Listeners were tested after a delay of approximately 40 s filled with other melodic material. All test items were transposed, and the shift was accomplished (or not) by controlling the harmonic context of the test item to either match (tonic-tonic, dominant-dominant) the scale position of the target, or to shift to another scale degree (tonic-dominant, dominant-tonic). Listeners were instructed to respond positively to all test items that were exact transpositions of the target, regardless of scale position. At least for moderately experienced listeners (ca. 6 years of music lessons), recognition of targets (vs. lures with changed pitches) was well above chance (.68) for test items not shifted along the scale, and fell to chance for shifted test items, despite the instructions. Thus, whether a melodic phrase is presented at the same relative scale position is important to delayed recognition, even when the test item is an exact transposition of the target.

It is noteworthy, though not surprising, that we observed no effects of musical training, suggesting that circadian synchrony has a more important influence on memory for musical phrases than do moderate levels of musical training. The lack of an effect of musical training is not surprising, given previously reported lack of training effects for various experimental tasks (see Bigand & Poulin-Charronnat, 2006, for a review) and given that, more specifically for the present short-term memory task, these improvement effects seldom depended on musical training in our earlier studies: in only one of the four experiments in Dowling et al.

(2001), and in only two of the eight experiments in Dowling and Tillmann (2014). When musical training effects did occur, it was usually in connection with some specific manipulation of the stimulus material. For example, when we contrasted naturally expressive minuet performances with more mechanical ones (Tillmann et al., 2013), musical training interacted with the type of musical performance. Whereas moderately trained listeners responded similarly to mechanical and to naturally expressive stimuli, untrained listeners showed a stronger decline in S false alarms across the delay with the expressive stimuli (much more like the experienced listeners) than with the mechanical stimuli. As the present stimuli are definitely naturally expressive (i.e., real performances of Ottmar Liebert), this feature probably aided the untrained participants in the present task. A second example of training effects in these studies is provided by Experiment 8 of Dowling et al. (2014), where we constructed S lures that differed from Ts *only* in the pitch level at which the melodic contour was bound to the scale, unlike the target phrases in the present experiment in which a number of other features such as the harmony and the pitch range differentiate S from T as well. In that case, moderately trained listeners showed T/S improvement over the delay, whereas untrained listeners did not. Presumably untrained listeners use the continued encoding of a variety of other cues to achieve the improvement effect when those other cues are available.

We were driven into a methodological innovation by the paucity of available excerpts for use as stimuli in the set of Ottmar Liebert albums we surveyed. We wanted to test four delay conditions, but with a total of 24 stimuli, we could only test two and still have enough trials in each condition to produce stable data points. Therefore, we chose the somewhat risky methodological strategy of having two groups perform the task at two delays each: 3 and 12 s, and 6 and 24 s. This strategy was risky because the resulting data might have turned out incoherent, for example, if the points for the 6-s delay condition for on-peak listeners had not fallen in between the points for the 3- and 12-s conditions. But for on-peak listeners, the data points for both area scores (Figure 2) and hits and false alarms (Figure 3) at the 6-s delay fell between the corresponding data points at the 3- and 12-s delays. This suggests that the T/S improvement effect is sufficiently robust that its effects operate largely independent of the testing context. That is, we can interrupt the process in midstream at 6 s, and obtain measurements entirely consistent with those we obtain by testing at 3 and 12 s.

Another methodological change from previous studies is the lack of structural continuity between the ongoing piece and the occurrence of the test item. In our earlier studies with minuets (Dowling & Tillmann, 2014), the test item generally occurred exactly where the composer had placed it in the piece. This ensured unbroken continuity between the target and the test item. It also ensured that whatever cues the composer had introduced to lead the listener to expect a particular phrase at that point in the piece would operate equally for all three types of test item. From the present results, it is clear that arbitrarily introducing the test item after a brief pause in the piece did not remove the memory improvement effect. Furthermore, in agreement with previous data (Dowling & Tillmann, 2014) showing the improvement effect when the perceived source, meter, and timbre remained coherent, the slight discontinuity that we introduced here between the end of the music excerpt (which had continued after the presentation of the target)

and the test item did not disrupt performance. This suggests that as long as the perceived source of the music remains stable, a minor structural discontinuity does not disrupt the ongoing processing of earlier material.

References

- Balzano, G. J., & Liesch, B. W. (1982). The role of chroma and scalestep in the recognition of musical intervals in and out of context. *Psychomusicology: A Journal of Research in Music Cognition*, 2, 3–31. <http://dx.doi.org/10.1037/h0094272>
- Bigand, E., & Poulin-Charronnat, B. (2006). Are we “experienced listeners”? A review of the musical capacities that do not depend on formal musical training. *Cognition*, 100, 100–130. <http://dx.doi.org/10.1016/j.cognition.2005.11.007>
- Bregman, A. S. (1990). *Auditory scene analysis*. Cambridge, MA: MIT Press.
- Dowling, W. J. (1978). Scale and contour: Two components of a theory of memory for melodies. *Psychological Review*, 85, 341–354. <http://dx.doi.org/10.1037/0033-295X.85.4.341>
- Dowling, W. J. (1984). Musical experience and tonal scales in the recognition of octave-scrambled melodies. *Psychomusicology: A Journal of Research in Music Cognition*, 4, 13–32. <http://dx.doi.org/10.1037/h0094206>
- Dowling, W. J. (1986). Context effects on melody recognition: Scale-step versus interval representations. *Music Perception*, 3, 281–296. <http://dx.doi.org/10.2307/40285338>
- Dowling, W. J. (1991). Pitch structure. In P. Howell, R. West, & I. Cross (Eds.), *Representing musical structure* (pp. 33–57). London, UK: Academic Press.
- Dowling, W. J., & Fujitani, D. S. (1971). Contour, interval, and pitch recognition in memory for melodies. *Journal of the Acoustical Society of America*, 49, 524–531. <http://dx.doi.org/10.1121/1.1912382>
- Dowling, W. J., Kwak, S., & Andrews, M. W. (1995). The time course of recognition of novel melodies. *Perception and Psychophysics*, 57, 136–149. <http://dx.doi.org/10.3758/BF03206500>
- Dowling, W. J., Lung, K. M.-T., & Herrbold, S. (1987). Aiming attention in pitch and time in the perception of interleaved melodies. *Perception and Psychophysics*, 41, 642–656. <http://dx.doi.org/10.3758/BF03210496>
- Dowling, W. J., & Tillmann, B. (2014). Memory improvement while hearing music: Effects of structural continuity on feature binding. *Music Perception*, 32, 11–32. <http://dx.doi.org/10.1525/mp.2014.32.1.11>
- Dowling, W. J., Tillmann, B., & Ayers, D. F. (2001). Memory and the experience of hearing music. *Music Perception*, 19, 249–276. <http://dx.doi.org/10.1525/mp.2001.19.2.249>
- Edworthy, J. (1985). Interval and contour in melody processing. *Music Perception*, 2, 375–388. <http://dx.doi.org/10.2307/40285305>
- Francès, R. (1988). *The perception of music* (W. J. Dowling, Trans.). Hillsdale, NJ: Erlbaum. (Original work published 1958)
- Hasher, L., Goldstein, D., & May, C. P. (2005). It’s about time: Circadian rhythms, memory, and aging. In C. Izawa & N. Ohta (Eds.) *Human learning & memory: Advances in theory and application: The 4th Tsukuba International Conference on Memory* (pp. 199–217). Mahwah, NJ: Lawrence Erlbaum Associates Publishers.
- Horne, J. A., & Östberg, O. (1976). A self-assessment questionnaire to determine morningness-eveningness in human circadian rhythms. *International Journal of Chronobiology*, 4, 97–110.
- Janata, P., Birk, J. L., Tillmann, B., & Bharucha, J. J. (2003). Online detection of tonal pop-out in modulating contexts. *Music Perception*, 20, 283–305. <http://dx.doi.org/10.1525/mp.2003.20.3.283>
- Koelsch, S. (2011). Towards a neural basis of processing musical semantics. *Physics of Life Reviews*, 8, 89–105.

- Koelsch, S., & Siebel, W. A. (2005). Towards a neural basis of music perception. *Trends in Cognitive Sciences*, 9, 578–584. <http://dx.doi.org/10.1016/j.tics.2005.10.001>
- Krumhansl, C. (1990). *Cognitive foundations of musical pitch*. New York, NY: Oxford University Press.
- Liebert, O. (1989). *Nouveau flamenco* [CD sound recording]. New York, NY: Higher Octave.
- Liebert, O. (1990). *Borrasca* [CD sound recording]. New York, NY: Higher Octave.
- Liebert, O. (1992). *Solo Para Ti* [CD sound recording]. New York, NY: Epic.
- Liebert, O. (1993). *The hours between night & day* [CD sound recording]. New York, NY: Epic.
- Liebert, O. (1995). *Viva!* [CD sound recording]. New York, NY: Epic.
- Liebert, O. (2001). *Little Wing* [CD sound recording]. New York, NY: Epic.
- Liebert, O. (2003). *Santa Fe Sessions* [CD sound recording]. New York, NY: Higher Octave.
- Schulze, K., Dowling, W. J., & Tillmann, B. (2012). Maintenance and Manipulation of tonal and atonal sequences in non-musician and musician listeners. *Music Perception*, 29, 255–267.
- Swets, J. A. (1973). The relative operating characteristic in psychology: A technique for isolating effects of response bias finds wide use in the study of perception and cognition. *Science*, 182, 990–1000. <http://dx.doi.org/10.1126/science.182.4116.990>
- Tillmann, B., & Dowling, W. J. (2007). Memory decreases for prose, but not for poetry. *Memory and Cognition*, 35, 628–639. <http://dx.doi.org/10.3758/BF03193301>
- Tillmann, B., Dowling, W. J., Lalitte, P., Molin, P., Schulze, K., Poulin-Charronnat, B., . . . Bigand, E. (2013). Influence of expressive versus mechanical musical performance on short-term memory for musical excerpts. *Music Perception*, 30, 419–425. <http://dx.doi.org/10.1525/mp.2013.30.4.419>
- Treisman, A. (2006). How the deployment of attention determines what we see. *Visual Cognition*, 14, 411–443. <http://dx.doi.org/10.1080/13506280500195250>
- Treisman, A., & Zhang, W. (2006). Location and binding in visual working memory. *Memory and Cognition*, 34, 1704–1719. <http://dx.doi.org/10.3758/BF03195932>
- Verde, M. E., 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. <http://dx.doi.org/10.3758/BF03208765>

Received August 4, 2014

Revision received January 11, 2016

Accepted January 11, 2016 ■

E-Mail Notification of Your Latest Issue Online!

Would you like to know when the next issue of your favorite APA journal will be available online? This service is now available to you. Sign up at <http://notify.apa.org/> and you will be notified by e-mail when issues of interest to you become available!