2005

Musical stem completion: Humming that note

J.A. Warker

Andrea Halpern
Bucknell University, ahalpern@bucknell.edu

Musical stem completion: Humming that note

JILL A. WARKER AND ANDREA R. HALPERN
Bucknell University

This study looked at how people store and retrieve tonal music explicitly and implicitly using a production task. Participants completed an implicit task (tune stem completion) followed by an explicit task (cued recall). The tasks were identical except for the instructions at test time. They listened to tunes and were then presented with tune stems from previously heard tunes and novel tunes. For the implicit task, they were asked to sing a note they thought would come next musically. For the explicit task, they were asked to sing the note they remembered as coming next. Experiment 1 found that people correctly completed significantly more old stems than new stems. Experiment 2 investigated the characteristics of music that fuel retrieval by varying a surface feature of the tune (same timbre or different timbre) from study to test and the encoding task (semantic or nonsemantic). Although we did not find that implicit and explicit memory for music were significantly dissociated for levels of processing, we did find that surface features of music affect semantic judgments and subsequent explicit retrieval.

In everyday life, people remember experiences and information using different methods of retrieval. Explicit memory, the conscious, intentional retrieval of previously learned material, is one of these methods. An example would be trying to recall a line from a movie. Another method of retrieval is implicit memory, or the retention of previously presented material without intentionally learning or retrieving that material. An example of this memory would be singing along to songs on the radio even though one has never studied the words. Most research on implicit memory has focused on the differences between explicit and implicit memory retrieval for verbal or nonverbal items presented visually. However, little research has investigated these differences in the auditory domain, particularly in the area of music. The purpose of this study was to learn more about the mechanisms of memory retrieval for auditory information with regard to music, including the characteristics of the stimuli that we focus on when retrieving that information.

The methods used to test whether one is exhibiting implicit or explicit memory differ from one another in significant ways. Implicit memory tasks involve automatic processing, whereas explicit memory tasks are seen as using conscious and attentional processing (Hayes & Hennessy, 1996).

AMERICAN JOURNAL OF PSYCHOLOGY
Winter 2005, Vol. 118, No. 4, pp. 567–585

In explicit tasks, participants are asked to recall previously presented information or to recognize previously presented information from a mix of old and novel information at test. Implicit memory tasks, on the other hand, usually involve a variation of priming, which occurs when an item is presented during study phase and that presentation facilitates unconscious access to the same information during a later test phase. The most commonly used task to test implicit memory is word stem completion. This task is often used to demonstrate implicit memory for written words (Brooks, Gibson, Friedman, & Yesavage, 1999; Lewandowsky, Dunn, & Kirsner, 1989; Ryan, Ostergaard, Norton, & Johnson, 2001). Participants read a list of words and are later given a list of word stems or three letters that start at least three separate words and are asked to complete the stem with the first word that comes to mind. Participants are most likely to complete the stems with the words they saw earlier, even though they are not intentionally trying to remember or produce these words and often fail to remember them when asked. This suggests that they are using their implicit memory to complete the stems. A comparable explicit memory test is cued recall. The procedure is identical except for the instructions. Participants study a word list and then are asked to recall words from the studied list to complete the word stems.

Many studies use word stem completion to demonstrate implicit retrieval for written words. A variation of the word stem completion test called picture fragment completion is also commonly used for pictorial stimuli. Participants look at pictures and are then presented with fragmented versions of those pictures as well as novel pictures. Participants needing fewer picture pieces to complete the old than new pictures are said to be exhibiting implicit memory (Rovee-Collier, Hayne, & Columbo, 2001). In an auditory stem completion test, lists of words are read aloud, and participants vocalize their completed word stems. Unbeknownst to them, participants often complete the word stems with the words they heard earlier (Pilotti, Gallo, & Roediger, 2000).

To date, little research has investigated implicit memory for nonverbal sounds, such as music. A few studies have demonstrated implicit memory for music using the mere exposure effect (Peretz, Gaudreau, & Bonnel, 1998; Gaudreau & Peretz, 1999; Halpern & O'Connor, 2000; Thompson, Balkwill, & Vernescu, 2000). Gaudreau and Peretz (1999) showed that tunes previously presented in the absence of explicit memory are liked better and judged as more familiar than unheard tunes, suggesting that we use our implicit memory to differentiate between tunes. Thompson et al. (2000) found that prior exposure to melodies influenced musical expectancies of subsequent notes. Participants rated previously heard melodies more highly than new melodies in which only the last note had been changed. A comparable experiment testing explicit memory was also conducted. Participants correctly recognized more old melodies than new melodies, but the results of the implicit memory and explicit memory tasks were uncorrelated. Peretz et al. (1998) found several variables that dissociated the implicit test (a mere exposure paradigm) and the explicit test (recognition).

Music shares some characteristics with other material studied in implicit memory paradigms. For instance, words and music vary in acoustic parameters such as phrasing, timbre, and pitch, and therefore retrieval for music might exhibit effects similar to those of studies with auditory verbal tasks. However, music is also different from previously studied domains in important ways. For instance, music could be considered more information intensive than words. One verbal phrase is composed of a few words, whereas one musical phrase is composed of many individual tones, intervals, and implied harmony. Also, unlike words or pictures, unfamiliar music has no connection to a mental lexicon (unfamiliar music typically is used in this type of study to avoid recoding by the tune's name). The representation must be built up de novo upon exposure.

Because the word and picture completion tasks have been so informative in the study of implicit memory processes, we wanted to create a stem completion task similar to those commonly used to test verbal and pictorial stimuli; this was the goal of Experiment 1. We called this production task tune stem completion. Participants listen to a tune and then hear a list of previously presented and novel tunes. However, each tune ends after a designated note, and the participants are asked to sing or hum the note they think would fit next musically. The explicit memory task was a cued recall task, which is nearly identical to the tune stem completion task with the exception of the instructions. The explicit instructions requested participants to complete the musical stem with the note that they remembered as coming next. The similar structure of the tasks allowed us to control the retrieval method participants used and to compare their performance on both tasks. We hypothesized that for both tasks participants would correctly complete more old stems than new stems. We also anticipated that performance in these two tasks would not be correlated.

If performance in these two tasks is uncorrelated, it is possible that implicit and explicit memory for music can be dissociated. A dissociation occurs when an independent variable affects performance on one type of task differently than it affects performance on another type of task. In order to determine whether such a dissociation exists, the retrieval intentionality criterion often is used (Schacter, Bowers, & Booker, 1989). This criterion has two stipulations. The first is that for both tasks, all study phase and external cues must be identical, but the instructions for the test phase of each task must differ, as in the tune stem completion task and the cued recall task. The second is that the experiment must include at least one independent variable that when manipulated affects performance on one task but not on the other, thus showing a dissociation between the retrieval processes (Rovee-Collier et al., 2001; Lewandowsky et al., 1989).

Experiment 2 manipulated two variables that we hypothesized would affect the musical stem completion and cued recall tasks in different ways. The first manipulated variable was encoding task, using tasks that should induce shallow or deep processing. Shallow processing occurs when one analyzes the surface characteristics of the information, including its physical and sensory features, such as pitch or color. Conversely, deep processing, also known as semantic encoding, occurs when one analyzes the meaning of the material, one's own past experiences with it, or one's personal judgments about it. Explicit memory is thought to be enhanced by encoding tasks stressing conceptual aspects of stimuli, such as semantic meanings, whereas implicit memory is thought to be enhanced by similarity in perceptual aspects of the stimuli from encoding to test (Lewandowsky et al., 1989).

Using implicit and explicit versions of a word fragment and word stem completion task, Roediger, Weldon, Stadler, and Riegler (1992) found that semantic encoding aided recall in the explicit version but had no effect on retrieval in the implicit version. These results further indicate that implicit memory relies more on perceptual characteristics, whereas explicit memory relies more on conceptual characteristics. Schacter and Church (1992) carried out a study using auditory verbal materials that illustrated this contrast. They manipulated voice gender in an auditory stem completion task and found that participants suffered reduced priming (i.e., implicit memory) when the voices differed between the study and test phases. They also used the same procedure in a cued recall test and found that performance was unaffected by the change in voices. This suggests that changes from study to test in the physical and surface characteristics of the stimuli are detrimental to implicit memory but have no effect on explicit memory. We should note that Peretz et al. (1998) did not find an effect of timbre change on ratings of tune pleasantness, although they used nameable, familiar timbres, unlike the nonspecific voice timbres used by Schacter and Church (1992).

Schacter and Church (1992) also looked at the effects of semantic and nonsemantic encoding on implicit memory and explicit memory. Half of the participants were instructed to judge the pleasantness of each word, and the other half were to judge how high or low the word's pitch was. The investigators found that participants who judged the pleasantness of the word recalled more information in the explicit memory task than participants who made pitch judgments. However, the different encoding conditions had no effect on participants' performance on the implicit memory task. These results further support the finding that deep processing facilitates explicit memory while having little or no effect on implicit memory. Few studies have looked at the effect of depth of processing on music. A study by Segalowitz, Cohen, Chan, and Prieru (2001) asked performance pianists to process songs shallowly or deeply under different elaboration methods (high and low). Using a cued recall task, they found that the pianists best recalled songs that were deeply processed with high elaboration, suggesting that deeply processing music leads to better explicit retrieval. Peretz et al. (1998) also found an effect of task elaboration on explicit recognition of melodies, although this was confined to familiar music.

Although studies have demonstrated how levels of processing and perceptual similarity affect implicit and explicit memory for spoken words differently, none have yet demonstrated these effects using tonal music in a stem completion task. Experiment 2 investigated this potential dissociation by varying the physical property of timbre of the melodies and by varying the encoding tasks between participants. As in Experiment 1, we expected that participants would perform better on old stems rather than new stems and that performance in the implicit memory and explicit memory task would not be correlated. In Experiment 2, we posited that implicit memory would be detrimentally affected when the physical properties of tunes changes from study to test, whereas explicit memory would remain unaffected. We also hypothesized that explicit memory would be enhanced when materials are semantically encoded, whereas implicit memory would be minimally affected. Overall, we expected that the results of these experiments would show that the retrieval processes for these two forms of memory are not only separate but also dissociated for tonal music because they are fueled by different characteristics of the stimuli.

EXPERIMENT 1

The purpose of Experiment 1 was to verify that implicit and explicit retrieval of music could be demonstrated using a vocal production task.

METHOD

Participants

Twenty-four undergraduates of Bucknell University enrolled in an introductory psychology course volunteered to participate in this experiment for class credit. Only students with 2 or more years of musical training who were willing to sing single notes were requested to volunteer. Students with prior musical training were requested to ensure that the sung notes would be sung in a recordable pitch. Participants had to respond to at least 10 out of 12 stems in each condition to be counted in the results; three participants were later eliminated for this reason.

Materials

Sixty tunes were selected from a collection of originally composed melodies previously created in our laboratory, some novel and some based on little-known folk tunes. The chosen tunes fulfilled several criteria. All tunes had to be tonal, nonverbal, and unfamiliar and contain at least one natural break in the first phrase. The tunes were 14.7 s long on average (range, 9–22 s). The beginning of the tune until the first natural break was called the musical stem. Stems were, on average, 6.2 s long (range, 4–10 s). Because this tune stem completion task was based on the word stem completion task, the musical stems had to be comparable to word stems, which begin several words other than the target word. To parallel a word stem, we ensured that for the note directly following each stem, the completion note, has at least one other note that could musically complete the stem. In general, the completion note was the final note in a musical phrase.

Five judges with musical backgrounds pilot tested the 60 tunes. They were asked to listen to 120 musical stems and rate how well the last note fit the stem musically on a scale of 1 (*not at all*) to 5 (*very much*). We wanted to make sure that at least one completion note other than the correct note would make for a natural-sounding phrase. The playlist consisted of 60 musical stems with the correct completion note and the same 60 musical stems with an alternate completion note. The scores were analyzed by comparing the scores for stems with the original ending with the scores for stems with the alternate ending. Tunes were eliminated for two reasons: a large difference between the rating of the correct stem and the alternate stem or low ratings for both stems. Of the 60 tunes that were tested, 12 were eliminated.

The materials used in Experiment 1 consisted of 48 tunes in their entirety and 48 stems, which stopped just before the first phrase ended (Figure 1). The tunes were separated into four groups of 12. The stems were separated into two groups that contained stems from the 12 previously heard tunes and stems from 12 novel tunes. All tunes and stems were counterbalanced over participants such that each tune and each stem occurred equally often as an old tune and stem and as a new tune and stem. The tunes and stems were synthesized in a grand piano timbre by Cakewalk software played through a Yamaha PSR500 keyboard. Experiments were controlled by a G3 Macintosh using SuperLab software.

Procedure

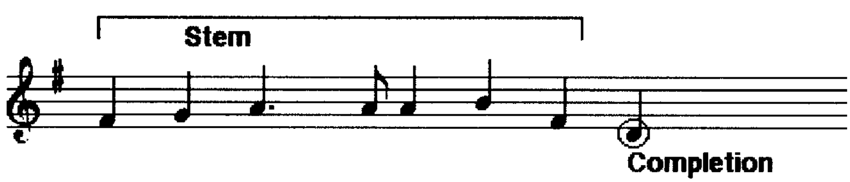

Figure 1. Example of a tune stem

The experiment used a 2 × 2 within-participant design. The two factors were old or new stems and implicit or explicit task. All participants were tested individually. Before beginning, participants completed a musical background questionnaire. Next they listened to 12 unfamiliar, tonal tunes at 2-s intervals. Participants were instructed, "Please listen to the following tunes," but were not told that they would be tested on the tunes later. Immediately afterward, they completed the implicit memory task. They heard 24 musical stems that stopped after a designated note; 12 had been previously heard, and 12 were novel. After each stem was played, they were asked to sing or hum the note that they thought would come next musically. The next task was the explicit memory task. Participants were told, "Please listen to another set of tunes." Participants were not instructed to remember the tunes for a later test because we wanted to keep all study and external cues the same between the two tasks. However, we assumed that participants would expect some kind of test after hearing the tunes because of the implicit task they had previously completed. Participants heard a different set of 12 unfamiliar, tonal tunes followed by a set of 24 previously heard and novel musical stems. After each stem was played, participants were asked to sing or hum the note that they remembered as coming next. Participants were not told that some of the musical stems they were hearing were novel. Rather, they were told that if they could not remember the note they were to guess the note that would come next. The notes sung by the participants in both tasks were identified using a Fender electric tuner, which displays the note that was sung to the nearest standard frequency in the equal tempered scale. Participants were then given a debriefing statement.

RESULTS

Scoring

The responses were scored on whether the sung note matched the original completion note. Notes were scored as correct if the sung note exactly matched the note originally chosen by the composer. Octave was discounted in the scoring. During the testing session, if a participant did not sing a recordable note for any one stem, that stem was not included in the analysis. For each participant, correctly completed stem scores were reported as the ratio of correct stems to the total number of answered stems and converted to percentages.

Old stems versus new stems

The results for the implicit and explicit tasks are displayed in Table 1. On average, participants correctly completed more old stems than new stems in both tasks. The difference in percentages for old and new stems was 20.6% ($SD = 22.8$) in the implicit task and 12.8% ($SD = 18.1$) in the explicit task. In both the implicit and the explicit task, participants completed significantly more old stems than new stems correctly (implicit task, $t(20) = 4.12$, $p = .0002$; explicit task, $t(20) = 3.24$, $p = .0021$). A correlation

Table 1. Mean percentages of tune stems completed correctly, Experiment 1

	Implicit memory		Explicit memory	
	Percentage correct	*SD*	Percentage correct	*SD*
Old stems	43.3	21.6	40.8	20.3
New stems	22.7	15.5	28.1	17.0
Old − New	20.6	22.9	12.8	18.1

was computed between each person's score on the implicit task and the explicit task. The correlation proved to be nonsignificant, $r(19) = .22$.

DISCUSSION

These results confirm that retrieval for tonal music can be demonstrated using a production task similar to those used in word stem and auditory stem completion studies. Previously, implicit memory for music had been found only using preference tasks. The results also demonstrate that results in the two retrieval tasks were not related, suggesting that whatever contributes to superior or inferior memory in one task does not do the same for the other task.

Testing memory for music using production, which is the way people commonly display memory for music in the real world, has its challenges in the laboratory. Many participants were hesitant when asked to hum or sing a note, even though they were forewarned about the task. Some data points were lost if the participant coughed, sneezed, or otherwise produced a hard-to-score pitch on a trial. It might be possible to elicit recall via a piano keyboard. However, in order for this method to work, participants would need to have prior knowledge of a keyboard, thus limiting experiment volunteers to those with a background in piano. Participants would also generate interference on each trial. All in all, vocal production seems to have been reasonably successful and provides a direct parallel to other stem completion procedures.

EXPERIMENT 2

This experiment investigated whether levels of processing and perceptual similarity of items at encoding and test affect implicit and explicit retrieval for music differently.

METHOD

Participants

Thirty-two undergraduates of Bucknell University enrolled in various psychology courses volunteered to participate in this experiment for class credit. None of the participants had taken part in Experiment 1. Only students with 2 or more years of musical training who were willing to sing single notes were requested to volunteer. Five participants were later excluded for low response levels. These participants responded to fewer than 13 out of 16 old stems or new stems, which is roughly equivalent to the criterion for eliminating participants in Experiment 1.

Materials

Twenty-four additional tunes matching the aforementioned criteria were compiled from the same collection of originally composed tunes and unfamiliar folk tunes used for Experiment 1. As in the pilot testing for Experiment 1, three judges with musical background rated how well the last note musically fit the stem on a scale of 1 (*not at all*) to 5 (*very much*). Stems with a difference between the original and alternate stem endings greater than 0.67 or with mean ratings less than 2.5 were dropped, leaving a total of 16 tunes. These 16 tunes were added to the original 48 tunes used in Experiment 1.

The final materials used for Experiment 2 consisted of 64 tunes in their entirety as well as 64 stems, which stopped after a designated note. The tunes were separated into four groups of 16. In each group, eight tunes were randomly assigned a flügelhorn timbre, and the remaining eight tunes were assigned a steel guitar timbre for the presentation phase. These timbres were chosen to be clearly distinguishable but not easily named. The stems were separated into two groups, which contained 16 previously heard stems and 16 novel stems. Participants heard one half of the stems in the same timbre as in the presentation phase and one half of the stems in a different timbre than at presentation. All stems were counterbalanced such that each stem occurred equally often in each timbre (flügelhorn and steel guitar) and in each condition (old and novel) over participants. The tunes and stems were synthesized by Cakewalk software played through a Yamaha PSR500 keyboard.

Procedure

The experiment had three within-participant factors (old or new stem, same or different timbre, and implicit or explicit task) and one between-participant factor (semantic or nonsemantic encoding). All participants were tested individually. Before beginning, participants completed a musical background questionnaire. In each task (implicit and explicit), participants listened to 16 tunes. Half of the participants were asked to judge the pleasantness of each tune, which we considered a semantic processing task, and half of the participants were asked to judge the regularity of the rhythm of each tune, which we considered a nonsemantic processing task. Tunes were rated on a 7-point scale from 1 (*least pleasant or least regular*) to 7 (*most pleasant or most regular*). Participants then heard 32 musical stems (half in same timbre and half in different timbre) that stopped after a designated note; half had been previously heard and half were novel.

Participants first completed the implicit memory task. After each stem was played, they were asked to sing or hum the note that they thought would come next musically. In the subsequent explicit memory task, participants listened to a different set of 16 unfamiliar tunes and were asked to perform the same encoding task that they performed during the first portion of the experiment. Participants then heard a set of previously heard and novel musical stems. After each stem was played, participants were asked to sing or hum the note that they remembered as coming next. The notes sung by the participants in both tasks were identified using a Fender electric tuner.

RESULTS

As an improvement over Experiment 1, individual items serving as new stems were reviewed to make sure that the correct completion note was not obvious. If the correct note was obvious, participants would not necessarily be exhibiting any type of memory retrieval for that particular stem. A stem was deemed to have an obvious completion note if the stem had at least 63.6% correct completions in the new stem condition (two standard deviations higher than the 27.3% mean completion for new tunes). Three stems created for Experiment 2 fit this criterion and were eliminated from all analyses. (None of the tunes used in Experiment 1 were found to have an obvious completion note, so no reanalysis was needed.) During the testing session, if a participant did not sing a recordable note for any one stem, that stem was not included in the analysis. Otherwise, response sheets were scored as in Experiment 1.

Participants in the semantic and nonsemantic encoding condition had similar music backgrounds (semantic, $M = 8$ years, $SD = 4$; nonsemantic, $M = 6.88$ years, $SD = 3.91$). The rating scores from the encoding tasks were briefly analyzed to make sure that participants were using the full range of the 7-point scale. For both tasks, participants rated the majority of songs from 3 to 6. The endpoints were used but not frequently. Participants in the semantic condition gave the tunes a mean rating of 4.32, $SD = 1.41$. Participants in the nonsemantic condition gave the tunes a mean rating of 4.44, $SD = 1.50$, *ns,* by a *t* test. Thus, the two groups appeared to be similar in background and overall approach to their encoding task.

The means and standard deviations for the implicit task and the explicit task are displayed in Table 2, which contains the percentage of correctly answered old and new stems.

Old stems versus new stems

As in Experiment 1, old stems were correctly completed significantly more often than new stems in both tasks. On average, participants cor-

Table 2. Mean percentages of tune stems completed correctly, Experiment 2

	Implicit memory		Explicit memory	
	Percentage correct	*SD*	Percentage correct	*SD*
Old stems	36.6	19.0	36.0	19.2
New stems	25.2	13.5	26.2	17.4
Old − New	11.5	15.3	9.8	12.3

rectly completed more old stems than new stems in both tasks (implicit task, $t(26) = 3.88$; explicit task, $t(26) = 4.13$, $p < .001$). The difference in percentages of old and new stems was 11.5%, $SD = 15.3$, for the implicit task and 9.8%, $SD = 12.3$, for the explicit task.

A 2 × 2 ANOVA was conducted using encoding condition as a between-participant variable and implicit or explicit task as the within-participant variable. The dependent measure was the difference between the percentage of correctly completed old stems and new stems. There was no significant main effect of encoding condition, $F(1, 25) = .10$, *ns,* nor of type of memory task, $F(1, 25) = .06$, *ns.* The predicted interaction between encoding condition and type of memory task was not significant, $F(1, 25) = .31$, *ns.* Judging the pleasantness of the tunes rather than the regularity of the rhythm did not result in greater memory performance for the explicit memory task than for the implicit memory task. Table 3 contains the mean and standard deviation results for old and new stems for the implicit and explicit memory tasks in the semantic encoding condition. Table 4 displays the results in the nonsemantic encoding condition.

Same versus different timbre

A 2 × (2 × 2) ANOVA was conducted using encoding task as a between-participant factor and the type of memory task and same or different timbre as the within-participant factors. The dependent measure was the percentage of correctly completed old stems. Old stems in the same timbre were not correctly completed significantly more often than old stems in a different timbre, $F(1, 25) = 1.17$, *ns.* The predicted interaction between old stems in a same or different timbre and the type of memory task was also not significant, $F(1, 25) = .37$, *ns.* See Tables 3 and 4.

Correlations

A correlation was computed between each person's score on the implicit task and the explicit task and was found to be significant, $r(25) = -.42$. This indicates that when implicit memory performance rose, explicit mem-

Table 3. Mean percentages of tune stems completed correctly in the semantic encoding condition, Experiment 2

	Implicit memory		Explicit memory	
	Percentage correct	*SD*	Percentage correct	*SD*
Old stems	34.3	12.3	37.4	19.5
Same timbre	36.7	12.0	38.0	23.1
Different timbre	28.1	23.9	36.7	20.4
Same − Different	8.6	30.2	1.3	18.6
New stems	24.9	12.9	26.6	18.9
Old − New	9.3	14.1	10.8	15.7

Table 4. Mean percentages of tune stems completed correctly in the nonsemantic encoding condition, Experiment 2

	Implicit memory		Explicit memory	
	Percentage correct	*SD*	Percentage correct	*SD*
Old stems	38.2	22.7	35.0	19.5
Same timbre	40.6	25.9	35.5	20.6
Different timbre	36.4	28.7	31.9	20.5
Same − Different	4.1	30.7	3.6	23.2
New stems	25.3	14.4	25.8	16.9
Old − New	12.9	16.4	9.1	9.9

ory performance declined. A scatter plot of the correlation data showed that two participants had outlying data points. A second correlation run without these data points was $r(23) = -.30$, a nonsignificant difference.

Using only Experiment 1 tunes

Because of the differences in results between Experiment 1 and 2, the data were rescored using only the tunes from Experiment 1. However, this analysis showed the same results found in Experiment 2.

Additional analysis of materials

The pleasantness ratings for the tunes in both timbres were analyzed. Although all tunes were counterbalanced and appeared equally often in each condition, tunes in the flügelhorn timbre were rated as significantly more pleasant, $M = 4.80$, $SD = 1.29$, than tunes in the steel guitar timbre, $M = 3.85$, $SD = 1.38$; $t(175) = 6.55$, $p < .001$. The rhythm regularity ratings for the tunes in both timbres were also analyzed. Tunes heard in the flügelhorn timbre were not rated as having a significantly more regular rhythmic pattern, $M = 4.51$, $SD = 1.54$, than tunes heard in the steel guitar timbre, $M = 4.38$, $SD = 1.46$; $t(255) = 1.04$, *ns*.

Because tunes in the flügelhorn timbre were found to be more pleasant, the old stems in same and different timbres were analyzed over both encoding conditions to see whether certain presentation–test timbre combinations were remembered better than others. The means and standard errors for all four timbre combinations in the implicit task can be seen in the top panel of Figure 2. In the implicit task, there was no significant difference among the timbre conditions, $F(1, 3) = .56$, *ns,* and no discernible trend. The means and standard errors for all four timbre conditions in the explicit task can be seen in the bottom panel of Figure 2. Although there was no significant overall difference among the four conditions in the explicit task, $F(1, 3) = 1.67$, *ns,* there was a significant linear trend, $F(1, 1) = 4.30$, $p = .048$. The means show that performance was best when

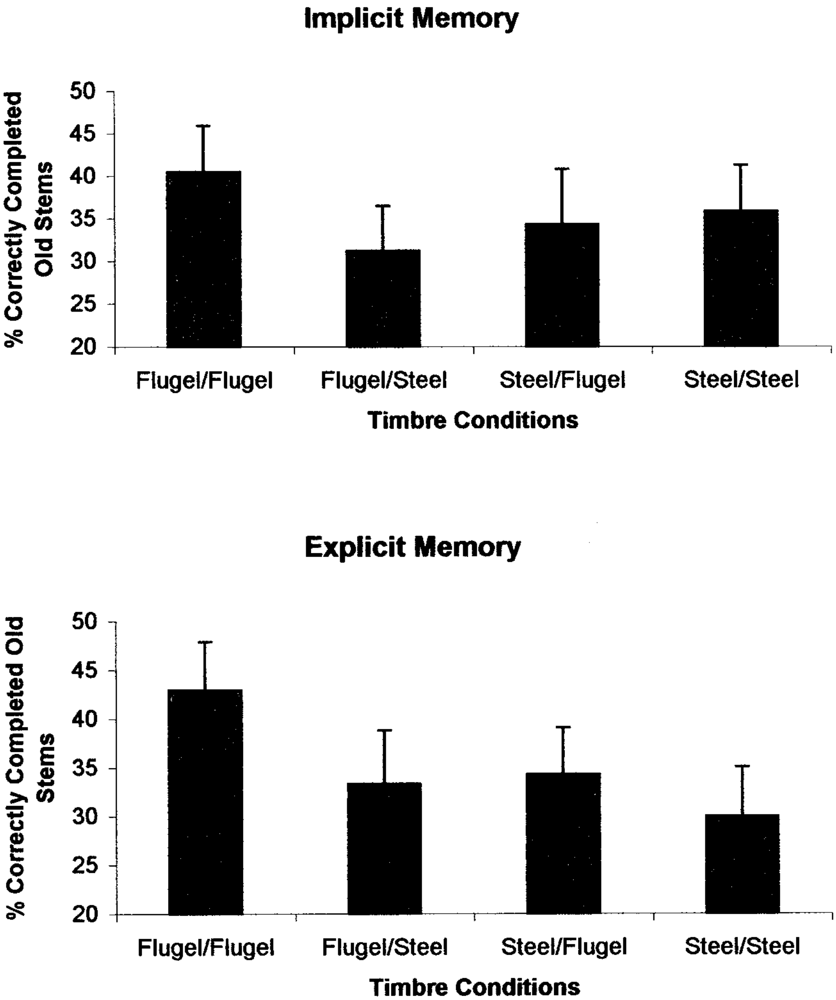

Figure 2. Means and standard errors for study timbre and test timbre combinations, Experiment 2

stems were in the preferred flügelhorn timbre at study and test and that performance declined as steel guitar replaced the preferred flügelhorn timbre.

DISCUSSION

In Experiment 2, old stems were correctly completed more often than new stems, replicating the stem completion finding in Experiment 1 and further confirming that tonal music can be implicitly retrieved using a production task. However, the negative correlation between performance on old stems and new stems was a surprising result. When the correlation was repeated without two participants with outlying data, the results coincided with those from Experiment 1.

Overall, memory performance on old items was worse in Experiment 2 than in Experiment 1 (36.3% correct vs. 42.1% correct). One reason for this discrepancy could be the greater memory load required of participants. Experiment 2 had participants encode 16 tunes and complete 32 stems in each task, compared with the 12 tunes and 24 stems used in Experiment 1. Another difference between the experiments was the encoding task. In Experiment 1, participants simply listened to the tunes, and in Experiment 2, participants made judgments about the tunes. It is possible that the levels of processing manipulation caused people to devote more attention to the task than to the music, especially because all the tunes were unfamiliar, as were the tasks. For instance, most people typically do not listen to new music with an ear to judging rhythmic regularity.

The results concerning the two manipulated variables, timbre and encoding, did not support our hypotheses. We expected performance in the implicit memory task to be lower when the tune stem was in a different timbre than previously heard at study, whereas performance in the explicit memory task would be unaffected. However, changing the physical characteristics of the stimuli affected neither implicit memory nor explicit memory even though previous studies on spoken words indicated that implicit memory performance should deteriorate when stimuli are spoken in different voices in study and test (Church & Schacter, 1994; Schacter & Church, 1992). One aspect that distinguishes voices is timbre, and so we had anticipated a similar result with our tunes. We have already noted that Peretz et al. (1998) did not find an effect of timbre change on implicit memory for either familiar or unfamiliar tunes, using more common and nameable instrument sounds. It is possible that other aspects of voice change, such as articulation patterns, might have increased the physical difference between voices in verbal studies compared with differences in synthesized instruments in the music studies. Alternatively, people might find voice quality to be a more separable aspect of spoken words than timbre is of melodies, and so music timbre might not function as a surface characteristic in music as much as it does in spoken words.

The second hypothesis was that participants in the semantic encoding condition would perform better on the explicit task than those in the nonsemantic condition, but encoding condition would have no effect on performance on the implicit task. The results indicate that encoding condition had no effect on performance in either task. There are several possible reasons why semantic encoding did not produce the predicted effect. The first is that the tunes did not vary enough in pleasantness. However, the analysis of ratings from the semantic encoding task shows that participants used the full range of the scale, although skewed toward the pleasant end. A future study could include more unpleasant tunes to make the task more salient and to make the task slightly easier, in case divided attention was an issue.

However, the additional analyses on the materials show another reason why encoding condition had no significant effect on performance. An analysis of the ratings in the semantic encoding condition found that participants rated tunes in the flügelhorn timbre as significantly more pleasant than those in the steel guitar timbre. This suggests that participants may have been rating the pleasantness of the timbre rather than the pleasantness of the melodic sequences. Thus, participants may not have been devoting their full attention to the tunes because they were distracted by a surface characteristic of the tune (i.e., the timbres) and therefore did not process the tunes at a deeper level. It is important to note that the timbre of the tunes did not affect rhythmic regularity ratings in the nonsemantic encoding condition, but it did affect pleasantness ratings in the semantic encoding condition. Although participants were distracted by the timbre of the tunes in the semantic encoding condition, they still produced an affective response. This suggests that the two encoding conditions were different from one another and that participants made perceptual judgments in the nonsemantic condition and affective judgments in the semantic condition, even if not in the requested manner.

The secondary analysis on the four timbre conditions for old tune stems showed that this preference for the flügelhorn affected performance in the explicit memory task. Participants performed best when tunes were in a flügelhorn timbre at study and test, less accurately but equally well in the two conditions when tunes switched timbres, and worst when tunes were in steel guitar at study and test. This suggests that timbre preference influences how people encode and explicitly recall music. It is important to note that timbre preference influenced explicit rather than implicit memory performance in the semantic encoding condition. This suggests that, as hypothesized, semantic encoding affects implicit memory and explicit memory differently, albeit in a more indirect way than we had anticipated.

GENERAL DISCUSSION

Experiments 1 and 2 found that the tune stem completion task is a valid production task to test implicit memory for music. The explicit cued recall task used was directly comparable to the tune stem completion task because only the instructions given during the test phase differed. Thus, we could assume that the only difference in performance between the implicit and explicit task was the result of the retrieval method the participant was using. The implicit memory results did not appear to be driven as a result of explicit memory processes. In Experiment 1, we found that performance on the two tasks was uncorrelated. In Experiment 2, we found a surprising inverse relationship between the two tasks when all data were included (this disappeared when two outliers were removed), which is difficult to explain in a principled way. Certainly, neither experiment suggested a contamination relationship between the two tasks.

In Experiment 2 we were unable to demonstrate a dissociation in explicit and implicit musical memory similar to that found in verbal memory. Contrary to our expectations, a salient physical change in the melodies between presentation and test did not impair the implicit more than the explicit measure, nor did a level of processing manipulation affect the explicit more than the implicit measure. In fact, neither manipulation was particularly successful. We can reject a floor effect explanation because memory was demonstrated in the percentage correct measure in both the implicit and explicit tasks. However, we also pointed out that memory was somewhat worse in Experiment 2 than in Experiment 1, probably because of the greater number of tunes presented. It is possible that the manipulations would have a greater impact under conditions of greater retention. This could be effected by, for instance, presenting all stimuli twice at presentation.

We mentioned earlier that the lack of an effect of timbre change in the implicit measure might result from the fact that voice change, used in auditory verbal stem completion, involves acoustic parameters other than timbre. Another reason timbre might not have affected performance in the predicted way is that in order for a timbre change to have negatively affected memory, tune identity (operationalized here as note completion) would have to have been bound with timbre in the memory representation. It may be the case that this binding operation is quite difficult in music. We used only two timbres and thought that reducing the set to only two would have made the binding requirements sufficiently easy. Thus, repeating these studies with more highly trained musicians, who presumably would be more adept at the binding operation, would be a relevant next step. Also, the timbres themselves were unfamiliar and hard to name, which also may have made encoding of the timbres ineffective (although they were clearly distinguishable).

We also failed to find a levels-of-processing (LOP) effect in explicit or implicit memory. We have few previous experiments on LOP effects in music with which to compare our results. Curiously, a database search on the terms *levels of processing and music, depth of processing and music, encoding* *and music, semantic encoding and music,* and *nonsemantic encoding and music* yielded only the study by Segalowitz et al. (2001) and the Peretz et al. (1998) study already discussed. Given the abundance of literature on LOP for verbal and pictorial information, it is hard to imagine that researchers have not tried to apply this approach to music. If they have, and bottom drawers are filled with unpublished musical LOP studies, it may simply be the case that music is encoded differently from other kinds of information or that "semantic" and "nonsemantic" distinctions must be reconsidered in this domain. Further evidence for this reconsideration can be found in event-related potential studies investigating violation of expectancies in music. Although violations of expected notes in a tune and expected words in a sentence both produce event-related potential changes, the locus and latency of these changes are somewhat different in the two domains (Besson & Faïeta, 1995). Because Peretz et al. (1998) found a direct influence of timbre on a recognition task and we found an indirect effect (in our case, timbre affected subjective pleasantness of melodies, which affected cued recall), it may be the case, as they suggest, that timbre forms part of a high-level object recognition process for melodies. If successfully encoded, timbre or affect might reasonably be expected to influence explicit rather than implicit memory. We also note that Peretz and colleagues used as their shallow encoding task classifying each instrument as piano or flute, which was of course the manipulation of surface structure they used. It is possible that LOP effects in familiar music may be revealed only when the manipulation is thus emphasized at encoding.

Our finding about the indirect effect of timbre pleasantness, not timbre identity per se, on memory success in the explicit task deserves another mention. Items encoded and retrieved in the more pleasant timbre (as defined by higher preference ratings for tunes in that timbre) were better remembered than tunes encoded and retrieved in the less pleasant timbre, with the intermediate case causing intermediate memory. This is notable because an encoding specificity argument would predict that having the same timbre at encoding and test should facilitate memory.[1] The fact that timbre influenced tune pleasantness judgments could mean that timbre, although not salient in memory, is quite salient during initial presentation of a tune. It could also mean that timbre interacts with other musical aspects to influence processing of musical characteristics of a tune needed for a pleasantness judgment. Crowder (1989) showed that a simple same versus different pitch judgment was affected by whether the two tones were in the same or different timbres (different timbres impaired detection of identical pitch). It is unclear at this time whether the timbre match effect reflects a true memory effect or perhaps a general facilitation of performance induced by simply being exposed to a more pleasant stimulus.

Notes

Jill A. Warker is now at the Department of Psychology, University of Illinois at Urbana–Champaign.

We thank James Bartlett for helpful discussions, Denise Chaykun for conducting pilot work, Taylor Harchak for scoring help, and our musical judges who rated the tune stems.

Portions of these experiments were presented at the meeting of the Psychonomic Society, Vancouver, BC, November 2003.

Correspondence about this article should be addressed to Jill A. Warker, Department of Psychology, University of Illinois, 603 E. Daniel St., Champaign, IL 61820 (e-mail: warker@s.psych.uiuc.edu). Received for publication May 24, 2004; revision received January 20, 2005.

1. We thank Bruno Repp for this observation.

References

Besson, M., & Faïeta, F. (1995). An event-related potential (ERP) study of musical expectancy: Comparison of musicians with nonmusicians. *Journal of Experimental Psychology: Human Perception and Performance, 21,* 1278–1296.

Brooks, J., Gibson, J., Friedman, L., & Yesavage, J. (1999). How to drastically reduce priming in word stem completion and still present the words. *Memory & Cognition, 27,* 208–219.

Church, B., & Schacter, D. (1994). Perceptual specificity of auditory priming: Implicit memory for voice intonation and fundamental frequency. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20,* 521–533.

Crowder, R. G. (1989). Imagery for musical timbre. *Journal of Experimental Psychology: Human Perception and Performance, 15,* 472–478.

Gaudreau, D., & Peretz, I. (1999). Implicit and explicit memory for music in old and young adults. *Brain and Cognition, 40,* 126–129.

Halpern, A. R., & O'Connor, M. G. (2000). Implicit memory for music in Alzheimer's disease. *Neuropsychology, 14,* 391–397.

Hayes, B. K., & Hennessy, R. (1996). The nature and development of nonverbal implicit memory. *Journal of Experimental Child Psychology, 63,* 22–43.

Lewandowsky, S., Dunn, J., & Kirsner, K. (1989). *Implicit memory: Theoretical issues.* Hillsdale, NJ: Erlbaum.

Peretz, I., Gaudreau, D., & Bonnel, A. (1998). Exposure effects on music preference and recognition. *Memory & Cognition, 26,* 884–902.

Pilotti, M., Gallo, D., & Roediger, H. L. III. (2000). Effects of hearing words, imaging words, and reading on auditory implicit and explicit memory tests. *Memory & Cognition, 28,* 1406–1418.

Roediger, H., Weldon, M., Stadler, M., & Riegler, G. (1992). Direct comparison of two implicit memory tests: Word fragments and word completions. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18,* 1251–1269.

Rovee-Collier, C., Hayne, H., & Colombo, M. (2001). *The development of implicit and explicit memory.* Philadelphia: John Benjamins.

Ryan, L., Ostergaard, A., Norton, L., & Johnson, J. (2001). Search and selection processes in implicit and explicit word-stem completion performance in young, middle-aged, and older adults. *Memory & Cognition, 29,* 678–690.

Schacter, D. L., Bowers, J., & Booker, J. (1989). Intention, awareness, and implicit memory: The retrieval intentionality criterion. In S. Lewandowsky, J. Dunn, & K. Kirsner (Eds.), *Implicit memory: Theoretical issues.* Hillsdale, NJ: Erlbaum.

Schacter, D. L., & Church, B. (1992). Auditory priming: Implicit and explicit memory for words and voices. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18,* 915–930.

Segalowitz, N., Cohen, P., Chan, A., & Prieru, T. (2001). Musical recall memory: Contributions of elaboration and depth of processing. *Psychology of Music, 29,* 139–148.

Thompson, W., Balkwill, L., & Vernescu, R. (2000). Expectancies generated by recent exposure to melodic sequences. *Memory & Cognition, 28,* 547–555.