Memory decreases for prose, but not for poetry

Barbara Tillmann
CNRS–UMR 5020, Université Lyon 1, Lyon, France

and

W. Jay Dowling
University of Texas at Dallas, Richardson, Texas

Memory for details of text generally declines relatively rapidly, whereas memory for propositional and context-based meanings is generally more resilient over time. In the present study, we investigated short-term memory for two kinds of verbal material: prose and poetry. Participants heard or read prose stories or poems in which a phrase near the start of the passage served as a target. The text continued, and after various delays, memory was tested with a repetition of the target (old verbatim; O), a paraphrased lure (P), or a lure in which the meaning was changed. For prose, memory for surface details (as measured by O/P discrimination) declined over time (Experiments 2–4), as was expected. For poetry, memory for surface details (O/P discrimination) did not decline with increasing delay (Experiments 1, 3, and 4). This lack of decline in memory for the surface details of poetry is discussed in relation to similar results previously observed for musical excerpts (Dowling, Tillmann, & Ayers, 2001), suggesting that a particular role is played by the temporal organization and rhythmic structure of poetry and music.

It seems to be a well-established fact that recall and recognition of items in short-term memory decline over time (e.g., Ebbinghaus, 1885). In addition, memory theories generally predict that additional material occurring during the delay between presentation of an item and its test interferes with memory performance (e.g., Baddeley, 1986, 1990; Waugh & Norman, 1965). This result has been reported not only for unconnected lists of words and sentences (e.g., Mehler & Carey, 1967), but also for coherent prose stories (Sachs, 1967). Sachs had participants listen to short prose stories and tested them during the ongoing presentation. A sentence near the beginning of the story was selected as the target. The story continued for varying delays, and a test phrase was presented that was identical to the original item (which we designate *old verbatim*; O), had its meaning changed (M), or had the same meaning as the original item but was altered in word order or verb form (a paraphrased lure; P). The participants had to discriminate the original target phrase from the changed foil phrases. Although immediate testing led to high performance overall, discrimination between Os and Ps dropped considerably with increasing delay, whereas O/M discrimination declined only slightly. This outcome reflects the loss of fine surface details over time, with relatively strong retention of semantic meaning. Gernsbacher (1985, 1990) has replicated and extended this data pattern with prose and picture stories.

Levels of Representation

An influential approach conceptualizing sentence and discourse memory distinguishes three levels in the mem-

ory representation of text (Kintsch, 1994; Kintsch, Welsch, Schmalhofer, & Zimny, 1990; van Dijk & Kintsch, 1983): the surface, verbatim form (consisting of the specific words and syntax); the textbase or propositional form (an abstract representation of the meaning in terms of the information conveyed in the text); and the situation model (an abstract characterization of the pragmatic situation that the text describes). Most studies have focused on the propositional text representation and the situation model (e.g., Fletcher & Chrysler, 1990; Schmalhofer & Glavanov, 1986), with only a few arguing for the importance of a surface level representation (Glanzer, Dorfman, & Kaplan, 1981). The three interrelated memory traces display distinct patterns of acquisition and retention in the time course of memory. Sachs's (1967) results illustrate the rapid decline in the strength of the surface representation and the longer retention of the propositional representation. Fletcher (1994) reviewed evidence that memory for the situation model can be even more resilient to forgetting than can memory for propositional structures (see also Kintsch et al., 1990, Experiment 1). Our study focused on the verbatim representation of text—in particular, for prose and poetry. In contrast to most studies on sentence and discourse memory, we were interested in examining the time course of memory for surface details in short-term memory (as measured by O/P discrimination).

Memory for Music

In contrast to the memory decline across delays generally observed for surface details in text or picture sto-

W. J. Dowling, jdowling@utdallas.edu

ries, studies on memory for music have reported a lack of decline and even an improvement in memory for details (Dowling & Tillmann, 2007; Dowling, Tillmann, & Ayers, 2001; see also Dowling, Kwak, & Andrews, 1995). Dowling et al. (2001) used a paradigm similar to that of Sachs (1967) to investigate memory for musical phrases embedded in ongoing musical pieces. On a given trial, listeners heard the beginning of a classical minuet, with the target phrase situated toward the start of the excerpt. The music continued, just as written by the composer and without interruption, for a short or a long delay (delays of 4–30 sec) until a high-pitched signal indicated the test phrase to follow. This test phrase was a repetition of the target (original item; O), a similar lure (S; sharing the melodic and rhythmic contour of O, but with pitch level or musical texture changed), or a different phrase (D). The listeners had to decide whether the test phrase was *exactly* the same as an earlier phrase they had heard in that same musical piece. Discrimination performance—in particular, the ability to discriminate targets from similar lures (O from S)—remained strong and even improved with increasing delay. This finding was surprising because it seemed to run counter to the usual result observed for prose stories—namely, that finer discrimination of changes in surface details (O from P) declines rapidly over time. For the music data, Dowling et al. (2001) proposed the hypothesis that the delay filled with music allows for continuous retrieval and encoding of the target item, which strengthens the memory trace of details (Tulving, 1983, 1984, 1986). After a long delay, the encoded information would then be different—richer and more closely linked to the target—from the information encoded after a short delay, thus improving the discrimination of targets from lures similar in surface details. This theoretical hypothesis emphasizes the importance of structural features linking the to-be-stored material and the contextual, supposedly interfering, material. The aim of the present study was to bridge the gap between the data patterns observed for music and prose by investigating memory for a material that shares the verbal character of prose while utilizing some of the structure of music in terms of rhythm and temporal organization—namely, poetry.

Memory for Poetry

The difference in the time course of memory for prose and music might be attributed to the difference between verbal and nonverbal (musical) materials—notably, the lack of semantic structures in music. Alternatively, the memory difference might be attributed to differences in the temporal organization of the material—particularly, its rhythmic structure. This second hypothesis focuses on the structural regularities in music (missing in prose) and suggests that the temporal, rhythmic structures in music, as well as its tonal structure, may have important effects on memory.

The aim of our present study was to test these two alternative explanations for the difference in memory performance between music and prose, focusing on differences in structural characteristics. Poetry appealed to us because it lies intermediate between prose and music in a space defined by these structural dimensions. On the one hand, poetry shares the syntactic structure and semantic meaning of prose. On the other hand, poetry, like music, often involves highly organized rhythmic structures and makes use of devices such as rhyme to mark the ends of phrases in a way analogous to the tonal cadences in music. That is, poetry makes use of rhythm and other sound features to bind phrases to each other and to the overall pattern, an aspect of stimulus organization it shares with music, and not with prose. Because of this overlap of poetry and prose in certain structural features (syntax and semantics) and with music in others (rhythmicity and temporal structure), our study compared memory for prose stories with memory for poetry. If the important factor explaining the difference between memory for prose and music is just that of verbal versus musical material (i.e., the presence or absence of semantic content), poetry should show a pattern of results that is the same as that for prose and different from that for music. If temporal structure is the important factor, poetry should produce results like those for music and different from those for prose.

EXPERIMENT 1

Experiment 1 adapted the experimental method previously used with musical material (Dowling et al., 2001) to poems. The target excerpts were either the very beginning of the poem (Lines 1 and 2) or the next two lines (Lines 3 and 4). Participants listened to the poem, and after a delay (short or long), the poem started again at the beginning (corresponding to the repeat in musical pieces) with either the same two lines (O) or a lure with the second line altered in surface structure (P) or meaning (M). The participants were required to make a *same/different* judgment indicating whether the test item was identical to a previously heard sentence or altered. Our study focused on the pattern of memory for detailed surface information (as captured by the O/P discrimination) across delay in order to test whether the pattern of the data for poetry would resemble that reported by Dowling et al. (2001) for music or that reported by Sachs (1967) for prose.

Method

Participants. Twenty-four undergraduate psychology students at the University of Texas at Dallas participated in groups of 5 to 8 persons for course credit. The answer sheet included a brief questionnaire of four questions on liking for, and writing of, poetry. The participants who answered above the middle of the scale on either "preferring poetry to other types of literature" or "number of poems written" were classified as "liking poetry." Fourteen participants were classified as "liking poetry," and the other 10 as "not liking poetry." Since this variable did not enter into any statistical effects, it will not be mentioned further.

Stimuli. Poems by A. E. Housman (1963) were selected because he wrote numerous poems that were similar in form and use of language and because we thought that, although unfamiliar, they would interest the participants. The poems were selected on the basis of length, metrical pattern, and rhyme scheme. They consisted of at least 12 three- or four-foot lines and followed a rhyme scheme of AABB or ABAB (Table 1). A different excerpt from a poem appeared on each trial. The excerpts usually began at the start of a poem, although in a few long poems, the excerpt consisted of the start of a section (defined in terms of content and versification). The

Table 1
Examples of Excerpts From Poems Structured As Trials in Experiment 1

Long Delay, Target in Lines 1 and 2, O Test Item

Why if 'tis dancing you would be,
There's brisker pipes than poetry.
Say, for what were hop-yards meant,
Or why was Burton built on Trent?
O many a peer of England brews
Livelier liquor than the muse,
And malt does more than Milton can
To justify God's ways to man.
Ale, man, ale's the stuff to drink
For fellows whom it hurts to think.

(beep)

Why if 'tis dancing you would be,
There's brisker pipes than poetry.

Short Delay, Target in Lines 3 and 4, P Test Item

The time you won your town the race
We chaired you through the market place.
Man and boy stood cheering by,
And home we brought you shoulder high.
To-day, the road all runners come,
Shoulder high we bring you home.
The time you won your town the race
We chaired you through the market place.

(beep)

Man and boy stood cheering by,
We brought you back home shoulder high.

Short Delay, Target in Lines 3 and 4, M Test Item

Others, I am not the first,
Have willed more mischief than they durst.
If in the breathless night, I too
Shiver now, 'tis nothing new.
More than I, if truth were told,
Have stood and sweated hot and cold.
Others, I am not the first,
Have willed more mischief than they durst.

(beep)

If in the breathless night, I too,
And miles around the wonder grew.

Note—O, old, verbatim target; P, paraphrased lure; M, meaning-changed lure.

experimental trials were defined by three conditions: (1) the position of the target lines, (2) the length of delay between the introduction of a target line and its test, and (3) the type of test item.

Target position (Lines 1 and 2 or Lines 3 and 4). On a given trial, the target consisted of either Lines 1 and 2 or Lines 3 and 4.

Testing delay (short or long). The number of lines intervening between the target item and the test item was four in the short-delay condition and eight in the long-delay condition, corresponding to approximately 32 and 64 syllables, respectively, or about 9.6 and 19.2 sec. More specifically, after the target, the excerpt continued for two or six lines. Then it continued for another two lines for the targets in Lines 1 and 2 and repeated Lines 1 and 2 for the targets in Lines 3 and 4. Then a brief, high-pitched beep was presented, which was followed by the test item. This pattern of delay and repetition paralleled the pattern we used in the music experiments (Dowling et al., 2001).

Test items (O, P, or M). O items repeated the target lines; P lures repeated the target lines but changed word order (sometimes involving a change in active or passive voice of the verb) in the second line of the target (i.e., in Line 2 or 4), while preserving rhythm and rhyme; M lures substituted different lines from poems not otherwise used in the experiment for the target, again preserving rhythm and rhyme.

The experiment consisted of 60 trials, with five items for each of the 12 target position × testing delay × test item combinations. The trials were divided into five blocks of 12 trials, each containing one instance of each condition combination, presented in random order. The poems were randomly assigned to the trial types, and two lists were prepared: Items tested at the short delay in List 1 were tested at the long delay in List 2, and vice versa. Approximately equal numbers of participants received each list. In four instances of List 1, the balance of trials within a block was disturbed; this was corrected in List 2.

For Experiments 1–3, the excerpts were read by the same female voice and were digitally recorded on one channel at CD quality (44.1-kHz sample rate, 16-bit resolution). The recordings were edited by means of Cakewalk Pro Audio 9, and beeps were inserted to alert the participants to the test items. O test items were exact replicas of the initially presented targets spliced into the appropriate places in the sequence. P and M lures consisted of newly recorded materials but were also spliced in the same manner as Os, so that any discontinuities in the auditory signal produced by splicing would occur equally in all the conditions. The stimuli were presented over loudspeakers at comfortable levels.

Procedure. The participants first filled out the questionnaire and were told that they would hear excerpts from poems. They were informed that near the end of each excerpt, the poem might begin to repeat and that a high-pitched beep would sound, indicating that the following lines constituted the test phrase. The participants were asked to judge whether or not they had heard *exactly* that test phrase before in that excerpt and to respond using a 6-point confidence-level scale (*very sure different*, *sure different*, *different*, *same*, *sure same*, and *very sure same*). The participants heard three sample trials, one each with O, P, and M test items, and the differences were explained to them. We emphasized to the participants that *any* difference was a sufficient reason to say "different" and that the response scale indicated their confidence in their judgment, and not a rating of the degree of difference.

Data analysis. Area under the memory-operating characteristic (MOC) provides an unbiased estimate of performance, where chance is .50 (Swets, 1973). Responses on the 6-point confidence-level scale were used to generate area scores with five criterion placements on the MOC. Area scores preserve more response information than do d' or A' and are uncorrelated with measures of response bias (unlike d'; see Dowling et al., 1995). Areas under the MOC were generated for comparisons between O items and P lures (O/P) and for comparisons between O items and M lures (O/M), for each condition and participant. In addition, proportions of hits for O items and false alarms for P and M lures were recorded for each condition and participant (with hit and false alarm rates being the proportion of *same* responses taken at the cut between responses 3 and 4 on the 6-point scale).

Results

Area scores. Area under the MOC was analyzed with a $2 \times 2 \times 2$ ANOVA, with item comparison (O/P or O/M), testing delay (short or long), and target position (Lines 1 and 2 or 3 and 4) as within-subjects factors. The main effect of item comparison was significant [$F(1,23) = 35.18$, $MS_e = 0.018$, $p < .0001$] and interacted with testing delay [$F(1,23) = 9.23$, $MS_e = 0.016$, $p < .01$; see Figure 1]. Performance was better overall for O/M comparisons than for O/P comparisons. Performance tended to decrease for O/M comparisons [$F(1,23) = 3.23$, $p = .09$], but not for O/P comparisons, which showed a difference in the opposite direction [even if not significant; $F(1,23) = 1.97$, $p = .17$]. In addition, the main effect of target position was significant [$F(1,23) = 9.10$, $MS_e = 0.03$, $p < .01$], with better performance for targets in Lines 1 and 2 (.78)
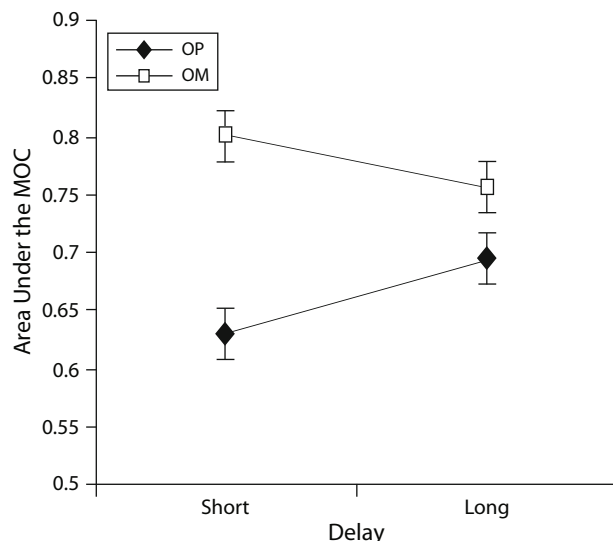
Figure 1. Area under the memory-operating characteristic (MOC) curve for O/P and O/M comparisons at short and long delays in Experiment 1 for poetry material. O/P indicates comparisons of old verbatim items and paraphrased lures; O/M indicates comparisons between old verbatim items and meaning-changed lures. Short and long delays corresponded to about 9.6 and 19.2 sec, respectively. Error bars indicate between-subjects standard errors.

than for those in Lines 3 and 4 (.68). No further effects were significant.

Hits and false alarms. Proportions of hits and false alarms (Table 2) were analyzed with a $3 \times 2 \times 2$ ANOVA, with test item (O, P, or M), testing delay (short or long), and target position (Lines 1 and 2 or 3 and 4) as within-subjects factors. A significant main effect of test item indicated that the participants discriminated successfully between original items and lures [$F(2,46) = 67.24$, $MS_e = 0.070$, $p < .0001$]. The interaction between test item and target position approached significance [$F(2,46) = 2.80$, $MS_e = 0.037$, $p = .07$] and corresponded to the effect of position for the area scores. No other effects were significant.

Discussion

Experiment 1 showed that memory for the fine details of short poetry excerpts did not decline over time. This stable performance pattern in the discrimination of P lures from O items contrasts with the performance decline obtained for prose stories (Sachs, 1967). For poetry material, the observed difference in means over time is more comparable to the pattern observed in memory for music (Dowling et al., 2001) than to that observed for prose. This outcome suggests that the difference in memory performance between Dowling et al. (2001) and Sachs was not simply due to differences between music and verbal materials. There remains, however, the methodological difference between Experiment 1 (which adapted the method of the music study) and Sachs's experiment. The most important difference in methodology, which might raise alternative explanations of the difference in results, was the fact that the target always occurred in the first four

lines of the poem and the first part of the test item was left intact, with only the second part being modified for the lures. This, plus the repetition of Lines 1 and 2 before a test of Lines 3 and 4, which might have provided additional cuing, could have influenced retrieval. Therefore, Experiment 2 applied the same experimental paradigm as that in Experiment 1 to prose material. If Experiment 2 showed the decline in memory performance over time that is typical of results with prose, this outcome would lead us to attribute the data pattern in Experiment 1 to similarities in temporal structures between music and poetry (such as rhythmic regularity and the use of closure devices, such as rhyme and tonality), and not to a methodological artifact arising from the structure of the trials.

EXPERIMENT 2

Method

Participants. Forty-six undergraduate psychology students at the University of Texas at Dallas participated in group sessions for course credit.

Stimuli. We used stories from the medieval Japanese *Tale of the Heike* (Kitagawa, 1977) because they lent themselves readily to paraphrasing into 16-syllable lines and it was highly probable that they would be unfamiliar to the participants. Their content echoed similar themes—love, death, war—to those of Housman's poetry. The excerpts were chosen from the starts of episodes in the story, usually the starts of chapters. A sample is shown in Table 3. Each excerpt was divided into sentences, and when necessary, the sentences were paraphrased or cut to fit into lines of 15–17 syllables, with a mean of 16 syllables. The majority of the sentences were 16 syllables in length. Occasionally, a long sentence occupied two lines, but such sentences never occurred as targets. It is important to note that the length of one line of prose corresponded to the length of two lines of poetry.

Trials were defined by two conditions: testing delay and test item. To avoid the possibility that the participants might anticipate targets in just one position (in case they noticed the regularity), the position of the target varied randomly between the first and the second lines of the excerpt; half the trials in each testing delay × test item condition placed the target in Line 1, and half in Line 2. Because of the small number of targets in either position, this was done simply for counterbalancing and was not tested in the design.

Testing delay (short or long). The test line followed the target line after two intervening lines in the short-delay condition and after six intervening lines in the long-delay condition. The two-line and six-line delays corresponded to approximately 32 and 96 syllables, respectively, or about 9.6 and 28.8 sec. The pattern of repetition characteristic of Experiment 1 was preserved: After the target, the excerpt continued for one line or five lines (in the short- and the long-delay conditions, respectively). For the targets in Line 1, the excerpt continued for another line, whereas for the targets in Line 2,

Table 2
Means of Hit Rates for Old Verbatim (O) Items and of False Alarm Rates for Paraphrased (P) and Meaning-Changed (M) Lures in Experiment 1, Presented As a Function of Target Position (Lines 1 and 2 vs. Lines 3 and 4) and Testing Delay (Short vs. Long)

Measure	Lines 1 and 2		Lines 3 and 4	
	Short	Long	Short	Long
Hits (O)	.65	.70	.63	.58
False alarms (P)	.40	.35	.39	.43
False alarms (M)	.15	.20	.19	.28

Table 3
Examples of Excerpts From Stories Structured As
Trials in Experiment 2

Long Delay, Target in Line 1, O Test Item

Shigekani, the chief archivist, was concerned about Jittei.
Shigekani was an exceptional man; capable in all matters.
One moonlit night Jittei had his lattice raised on the south side.
He was chanting poems in the moonlight when Shigekani came by.
Who is there, he asked. It is Shigekani, my lord, was the reply.
On what business have you come to see me? lord Jittei asked him.
I come in meditation, washed clean in the beauty of the moonlight.
(beep)
Shigekani, the chief archivist, was concerned about Jittei.

Short Delay, Target in Line 2, P Test Item

Why should I be burdened with an ugly woman whom I dislike?
If I chose the one I love, I would be thought a disloyal son.
Now must I open my eyes to find a way to Buddhahood.
Why should I be burdened with an ugly woman whom I dislike?
(beep)
They would think me a disloyal son if I chose the one I love.

Long Delay, Target in Line 2, M Test Item

Her maids teased her and said, Where did that moon come from? Tell us, please.
But Tadamori's heart was touched by the poem that she penned in reply.
And she became the mother of Tadamori's son, Tadanori.
Tadamori became lord high marshal and died at fifty-eight.
His oldest son Kiyomori assumed the duties of his household.
Then the notorious Yorinaga raised the standard of revolt.
Kiyomori stood steadfast by the emperor and was promoted.
Her maids teased her and said, Where did that moon come from? Tell us, please.
(beep)
But the lady's heart was touched by the poem Tadamori penned in reply.

Note—O, old, verbatim target; P, paraphrased lure; M, meaning-changed lure.

the excerpt repeated Line 1. Then a brief, high-pitched beep was presented and followed by the test item. In the second example in Table 3, the target is Line 2. For this short-delay trial, two lines intervened between target and test: one line continuing the story and the next line repeating Line 1. Note the parallel structure with the second example in Table 1, which was a short-delay P item of poetry with the target in Line 2.

Test items (O targets, P lures, M lures). O targets repeated the target lines; P lures repeated the target lines, but with changed word order that left the meaning intact; and M lures repeated target lines, but with changes in wording that altered the meaning.

The experiment consisted of 24 trials, including four items for each of the 2 testing delay × 3 test item combinations. The excerpts were randomly assigned to trial types, and trials were arranged in two blocks of 12 trials (containing two items of each combination), presented in random order.

Procedure. The procedure was exactly the same as that in Experiment 1, except that the participants were told that they would hear excerpts from medieval Japanese stories and were not asked about their liking for poetry.

Results

Area scores. Area under the MOC (Figure 2) was analyzed with a 2 × 2 ANOVA, with item comparison (O/P or O/M) and testing delay (short or long) as within-subjects factors. The main effects of item comparison and of testing delay were significant [$F(1,45) = 24.43$, $MS_e = 0.015$, $p < .0001$, and $F(1,45) = 31.09$, $MS_e = 0.047$, $p < .0001$, respectively]. Performance was better overall for O/M comparisons than for O/P comparisons and decreased with increasing delay. The item comparison × testing delay interaction approached significance [$F(1,45) = 3.57$, $MS_e = 0.015$, $p < .07$]. In contrast to Experiment 1, performance declined for both O/P and O/M comparisons, with stronger decline for O/P [$F(1,45) = 28.03$, $R^2 = .20$, $p < .0001$] than for O/M [$F(1,45) = 17.01$, $R^2 = .08$, $p < .001$].

Hits and false alarms. Proportions of hits and false alarms (Table 4) were analyzed with a 3 × 2 ANOVA with test item (O, P, or M) and testing delay (short or long) as within-subjects factors. The main effect of test item was significant [$F(2,90) = 63.55$, $MS_e = 0.06$, $p < .0001$], indicating that the participants successfully discriminated between original items and lures. The interaction between test item and testing delay (parallel to the effect of delay for the area scores) was also significant [$F(2,90) = 11.74$, $MS_e = 0.05$, $p < .05$]. Hits for O items decreased with delay, whereas false alarm rates for P and M lures increased, with a particularly strong increase for P lures. No other effects were observed.

Discussion

In Experiment 2, we applied to prose material the same experimental method as that used for poetry material in Experiment 1. In contrast to the results with poetry, performance declined over time for O/P comparisons. This outcome is in agreement with Sachs's (1967) results for prose stories, even though our experimental method was different from hers. In Experiment 3, we compared prose and poetry directly, using a method closely modeled on that of Sachs.
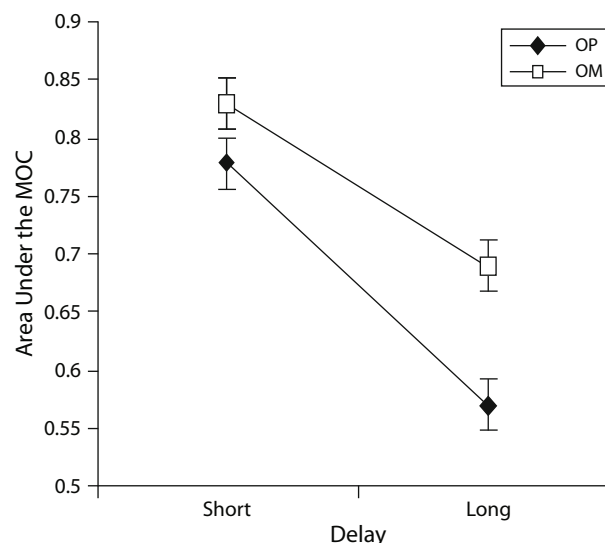
Figure 2. Area under the memory-operating characteristic (MOC) curve for O/P and O/M comparisons at short and long delays in Experiment 2 for prose material. O/P indicates comparisons of old verbatim items and paraphrased lures; O/M indicates comparisons between old verbatim items and meaning-changed lures. Short and long delays corresponded to about 9.6 and 28.8 sec, respectively. Error bars indicate between-subjects standard errors.

Table 4
Means of Hit Rates for Old Verbatim (O) Items and of False Alarm Rates for Paraphrased (P) and Meaning-Changed (M) Lures, Presented As a Function of Testing Delay (Short vs. Long) for Experiments 2–4

| | Hits (O Items) | | False Alarms | | | |
| | | | (P Items) | | (M Items) | |
Materials	Short	Long	Short	Long	Short	Long
Experiment 2						
Prose	.72	.56	.30	.44	.22	.25
Experiment 3						
Prose	.81	.68	.26	.38	.29	.24
Poetry	.85	.67	.55	.23	.33	.45
Experiment 4						
Prose	.80	.66	.48	.48	–	–
Poetry	.66	.70	.55	.51	–	–

EXPERIMENT 3

In Experiment 3, we studied memory for prose and poetry with an experimental method that preserved the main features of Sachs's method: Targets occurred in variable positions, and no repetition of material preceded the test item. Without altering the essentials of her design, we adapted Sachs's (1967) method with slight alterations in the set of testing delays, in the definition of test items, and in our use of a different set of prose stories. Only two (instead of three) testing delays were used, omitting her longest delay and substituting for her immediate test a delay involving one interpolated line. This gave a delay of approximately 5 sec, comparable to the delays in Experiments 1 and 2. As in Experiment 2, only three test item types (instead of Sachs's four) were used: O items (exact repetitions, as in Sachs), M lures (corresponding to Sachs's semantic-alteration lures), and P lures involving changes of word order that did not alter meaning (collapsing Sachs's two categories of surface alterations, because of the similar results in her experiment). We adapted this method to prose and poetry material. For prose, we predicted a performance decrease with delay for surface details (O/P discrimination), just as in Experiment 2 and in Sachs's experiment. For poetry, we predicted no decline in O/P performance, as in Experiment 1.

Method

Participants. Ninety-five undergraduate psychology students at the University of Texas at Dallas participated in group sessions for course credit: 43 with the prose material and 52 with the poetry material. In the poetry group, 28 participants professed to like poetry, and 24 not. (Seven additional participants performed at no better than chance in terms of hits and false alarms for both test item types in the long-delay condition and were discarded from the analysis.) In the prose group, we administered the poetry questionnaire to the last 22 participants, since it occurred to us that the similarity of the two groups with regard to their liking for poetry should be documented: Thirteen professed a liking for poetry versus nine who did not. This proportion did not differ by χ^2 from that observed for the poetry group. Since this variable did not enter into any statistical effects, it will not be mentioned further.

Stimuli. For the prose material, we used the same stories as those in Experiment 2 (see Table 5). Following Sachs (1967), target lines were randomly selected from Lines 3–12 (inclusive) of the excerpt, with a distribution of possible positions with a mean of six lines. For the poetry material, we made a new selection of poems by A. E. Hous-

man (1963). Target locations were chosen randomly so that targets occurred equally often in Lines 3 and 4, 5 and 6, 7 and 8, or 9 and 10.

Testing delay (short or long). In both conditions, the excerpt continued after the target for about 4.8 or 28.8 sec in the short- and long-delay conditions, respectively. The delays corresponded to 1 line and 6 lines for the prose and to 2 and 12 lines in the poetry (approximately 16 and 96 syllables, respectively).

Test item (O targets, P lures, M lures). O items repeated the target lines; P lures involved a change of wording that did not alter meaning; and M lures contained a semantic change. For poetry, alterations in P and M test items involved changes in both lines of the target (unlike in Experiment 1).

For prose and poetry, respectively, six conditions were defined by 2 testing delays and 3 test items; each condition was represented by 4 trials, leading to 24 trials for each participant. For the poetry material, two counterbalanced lists were constructed so that items tested at the short delay in List 1 were tested at the long delay in List 2 (and vice versa) and items tested with one test item type in one list were tested with another test item type in the other list, determined randomly. An additional control feature was introduced

Table 5
Examples of Excerpts From Stories Structured As Trials in Experiment 3

Long Delay, Target in Line 3, O Test Item

Lord Jittei, though worthy, was passed over for a generalship.
He secluded himself and decided to enter the priesthood.
Shigekani, the chief archivist, was concerned about Jittei.
Shigekani was an exceptional man; capable in all matters.
One moonlit night Jittei had his lattice raised on the south side.
He was chanting poems in the moonlight when Shigekani came by.
Who is there, he asked. It is Shigekani, my lord, was the reply.
On what business have you come to see me? lord Jittei asked him.
I come in meditation, washed clean in the beauty of the moonlight.
(beep)
Shigekani, the chief archivist, was concerned about Jittei.

Short Delay, Target in Line 10, P Test Item

Tokiyori was the son of the Chief Guard of the Imperial Gate.
When he was thirteen Tokiyori visited the Archives Office.
There he happened to meet a low-ranking maid named Yokobue.
Tokiyori was immediately captivated by her charms.
His father became angry and said, I have other plans for you.
I shall find you a good match from a family influential at court.
Follow your father's advice and stop loving such a lowly maid.
Tokiyori replied, our life is as fleeting as a spark from flint.
Why should I be burdened with an ugly woman whom I dislike?
If I chose the one I love, I would be thought a disloyal son.
Now must I open my eyes to find a way to Buddhahood.
(beep)
They would think me a disloyal son if I chose the one I love.

Long Delay, Target in Line 5, M Test Item

The woman Tadamori loved best was in the Emperor's entourage.
While at court Tadamori found frequent occasions to visit her.
One day he left behind a fan with a moon painted in its corner.
Her maids teased her and said, Where did that moon come from? Tell us, please.
But Tadamori's heart was touched by the poem that she penned in reply.
And she became the mother of Tadamori's son, Tadanori.
Tadamori became lord high marshal and died at fifty-eight.
His oldest son Kiyomori assumed the duties of his household.
Then the notorious Yorinaga raised the standard of revolt.
Kiyomori stood steadfast by the emperor and was promoted.
When Nobuyori revolted Kiyomori helped put him down.
(beep)
But the lady's heart was touched by the poem Tadamori penned in reply.

Note—O, old, verbatim target; P, paraphrased lure; M, meaning-changed lure.

for poetry in Experiment 3 and for both poetry and prose in Experiment 4—namely, *reversed-text* control trials (Fletcher, 1992). This control addresses the potential criticism that since P and M test items were altered from the author's original text, better-than-chance performance might be achieved by discriminating the original text from the experimenter-produced imitations. (This is especially important for poetry, where, presumably, the author has carefully chosen the exact word order.) Therefore, for O and P trials, a P item was constructed corresponding to the target, and on half the trials of each type (randomly selected), the P item was substituted for the original target at the beginning of the trial. The substituted P item was then repeated in the test position on O trials, whereas the O item appeared in the test position on P trials. For half the M trials (randomly selected), we exchanged targets and test items (O and M items).

Procedure. The procedure was as described in Experiment 2 for prose and in Experiment 1 for poetry.

Results

Area scores. Area under the MOC (Figure 3) was analyzed with a $2 \times 2 \times 2$ ANOVA, with item comparison (O/P or O/M) and testing delay (short or long) as within-subjects factors and material (prose or poetry) as a between-subjects factor. The three-way item comparison $\times$ testing delay $\times$ material interaction was significant [$F(1,93) = 43.71$, $MS_e = 0.018$, $p < .0001$]. For prose, performance decreased with delay to a stronger extent for O/P comparisons [$F(1,71) = 14.36$, $p < .001$] than for O/M comparisons ($F < 1$, n.s.). For poetry, performance decreased strongly with delay for O/M comparisons [$F(1,71) = 29.67$, $p < .0001$] but improved slightly [although not significantly; $F(1,71) = 1.22$, $p = .27$] for O/P comparisons. In addition, the main effects of testing delay [$F(1,93) = 19.57$, $MS_e = 0.036$, $p < .0001$] and item comparison [$F(1,93) = 6.37$, $MS_e = 0.014$, $p < .05$] and the interaction between test item and material [$F(1,93) = 6.17$, $MS_e = 0.014$, $p < .05$] were significant.

Hits and false alarms. Proportions of hits and false alarms (Table 4) were analyzed with a $3 \times 2 \times 2$ ANOVA with test item (O, P, or M) and testing delay (short or long) as within-subjects factors and material (prose or poetry) as a between-subjects factor. The main effect of test item was significant [$F(2,186) = 175.62$, $MS_e = 0.054$, $p < .0001$], indicating that the participants successfully discriminated between O items and lures. The interaction of test item and testing delay [$F(2,186) = 4.71$, $MS_e = 0.046$, $p < .05$] was modified by the material factor, as expressed in a three-way test item $\times$ testing delay $\times$ material interaction [$F(2,186) = 25.24$, $MS_e = 0.046$, $p < .0001$]. With increasing delay, hit rates decreased for poetry [$F(1,71) = 22.82$, $p < .0001$] and prose [$F(1,71) = 4.99$, $p < .05$], whereas false alarms for P items decreased for poetry [$F(1,71) = 65.02$, $p < .0001$] but tended to increase for prose [$F(1,71) = 3.69$, $p < .07$]. For M items, false alarm rates increased for poetry [$F(1,71) = 7.31$, $p < .01$] but remained rather stable for prose ($F < 1$, n.s.). For poetry alone, the interaction between test item and testing delay [$F(2,144) = 14.95$, $MS_e = 0.053$, $p < .01$] and the main effects of testing delay [$F(1,93) = 12.73$, $MS_e = 0.066$, $p < .001$] and of material [$F(1,93) = 8.41$, $MS_e = 0.066$, $p < .01$] were significant.

Discussion

Experiment 3 adapted Sachs's experimental method to both prose and poetry—in particular, utilizing random selection of target position and omitting the repetition of the lines preceding the target that characterized Experiments 1 and 2. For prose, the results replicated Sachs's result, as well as the results of Experiment 2: O/P discrimination declined between short and long delays. However, for poetry, the outcome replicated the pattern in Experiment 1, with no decline in O/P discrimination. For poetry, O/P performance in Experiment 3 was due mainly to an improved ability to reject P lures over time. The fact that this pattern was observed despite the uncertainty of target
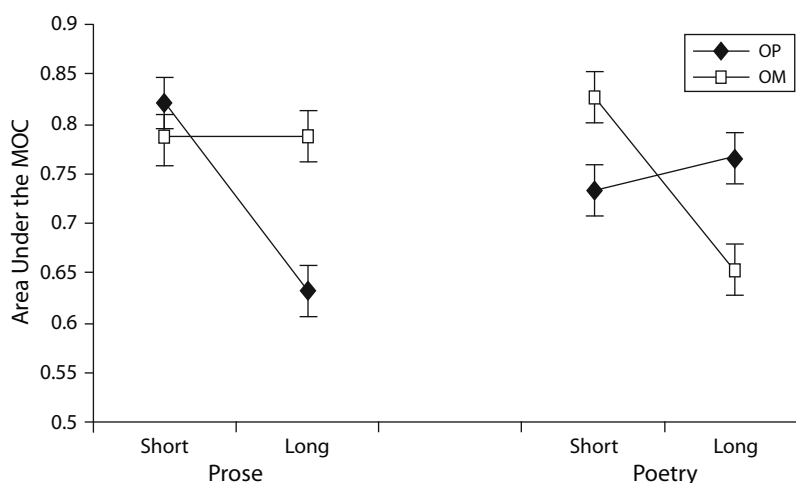

Figure 3. Area under the memory-operating characteristic (MOC) curve for O/P and O/M comparisons at short and long delays in Experiment 3 for prose and poetry material. O/P indicates comparisons of old verbatim items and paraphrased lures; O/M indicates comparisons between old verbatim items and meaning-changed lures. Short and long delays corresponded to about 4.8 and 28.8 sec, respectively. Error bars indicate between-subjects standard errors.

location supports the notion of rather automatic, unconscious processes being involved in O/P discrimination.

EXPERIMENT 4

Experiments 1–3 showed differences between poetry and prose in memory for surface details over time (as measured by O/P discrimination). O/P performance declined over time for prose, but not for poetry. In these experiments, the same speaker recorded both the prose and the poetry. With the goals of excluding possible differences in oral production between prose and poetry and of extending these results to the works of another poet, Experiment 4 used visual presentation of the material and drew the poetry from Edward Fitzgerald's translation of the *Rubaiyat of Omar Khayyam*. Experiment 4 tested memory for poetry and prose only in terms of O/P discrimination, in order to focus on memory for surface details.

Method

Participants. Forty-three undergraduate psychology students at the University of Texas at Dallas participated in individual sessions for course credit: 23 with prose and 20 with poetry. Each participant completed a brief questionnaire at the top of the answer sheet, as described in Experiment 1. Twenty-two participants were classified as "liking poetry," and 21 as "not liking poetry." As in Experiments 1 and 3, since this variable did not enter into any statistical effects, it will not be mentioned further.

Stimuli. We used the *Rubaiyat of Omar Khayyam* (translated by Edward Fitzgerald, 5th edition; Williams, 1947, pp. 264–278). These poems were similar in rhythm and rhyme to those of Housman (1963) used in Experiments 1 and 3. Fitzgerald's translation provides a continuous sequence of 404 lines, which forms a relatively coherent whole when presented sequentially over a series of trials, as was done in the present experiment. As with the Housman poems, our aim was to select poems that the readers would find exciting and interesting, dealing with love, death, and the human condition. The poems are organized into quatrains of five-foot lines in iambic pentameter, rhymed AABA (see Table 6). The prose material was drawn from the *Tale of the Heike* (Kitagawa, 1977), as in Experiments 2 and 3. Each of the lines of prose, as is illustrated in Table 3, was divided in half for presentation on successive slides.

Stories and poems were presented on the computer screen one line at a time in sequential order at a rate of 4 sec/line, using PowerPoint presentation software. Target lines were chosen randomly so as to occur one to six lines after the previous test position (or the beginning of the sequence). Delays between a target line and its test involved either two or six intervening lines. Immediately preceding a test line, the word TEST appeared, centered on the screen, for 1 sec. Thus, the delay between target and test was either 9 sec (short) or 25 sec (long), or about 16–20 syllables and 56–60 syllables, respectively. The test line was presented for 6 sec, during which the participant wrote the response on the answer sheet, using the 6-point confidence-level scale described above. A warning signal (i.e., an x in the middle of the screen) preceded the next item by 2 sec.

For poetry and prose, respectively, 32 trials were defined by testing delay (short or long) and test item (O or P), and each trial type was represented by eight instances. Half of the test items involved the reversed-text control (Fletcher, 1992) described for Experiment 3. Trials were presented in a random order determined by two random permutations of 16 digits, with each trial type equally represented in the first and second halves of the list. The slides were displayed in 32-point Times New Roman, with the TEST and x warning slides in 24-point Copperplate Gothic. The participant's eyes were approximately 0.6 m from the screen.

Procedure. The participants first filled out the questionnaire and were told that they would view a series of poems (the *Rubaiyat of*

Table 6
Examples of Excerpts From Poems Structured As Trials in Experiment 4

Short Delay, Target in Line 6, P Test Item
(Reversed-Control Condition; Lines Represent Slides in Slide Show)

Wake! For the Sun, who scattered into flight
The Stars before him from the Field of Night,
Drives Night along with them from Heav'n, and strikes
The Sultan's Turret with a Shaft of Light.
Before the phantom of False morning died,
A Voice methought within the Tavern cried,
"When all the Temple is prepared within,
Why nods the drowsy Worshipper outside?"

TEST

Methought a voice within the Tavern cried,

x

Long Delay, Target in Line 4, O Test Item
(Poem Continues From First Example)

And those who, as the Cock crew, stood before
The Tavern shouted—"Open then the door!
You know how little while we have to stay,
And, once departed, may return no more."
Now the New Year reviving old Desires,
The thoughtful Soul to Solitude retires,
Where the White Hand of Moses on the Bough
Puts out, and Jesus from the Ground suspires.
Iram indeed is gone with all his Rose,
And Jamshyd's Sev'n-ring'd Cup where no one knows;

TEST

And, once departed, may return no more."

x

But still a Ruby gushes from the Vine,

etc.

Note—P, paraphrased lure; O, old, verbatim target.

Omar Khayyam) or some story excerpts (from the *Tale of the Heike*) on the computer screen. They were told they would see a series of lines from the poems or stories on the screen, each presented for about 4 sec. From time to time, there would be a test trial, introduced by the word TEST in the center of the screen. Their task was to decide whether the line presented on the test trial had occurred in *exactly* the same form earlier in the experiment, during the preceding 10 or 12 lines. They were required to write their responses in successive blanks on the answer sheet, using the confidence-level scale printed there.

Results

Area scores. Area scores were analyzed with a 2 × 2 ANOVA, with testing delay (short or long) as a within-subjects factor and material (poetry or prose) as a between-subjects factor. Only the testing delay × material interaction was significant [$F(1,41) = 4.13$, $MS_e = 0.020$, $p < .05$]. As is shown in Figure 4, prose performance decreased over the long delay interval [$F(1,41) = 5.41$, $MS_e = 0.02$, $p < .05$], whereas poetry performance did not decline.

Hits and false alarms. Proportions of hits and false alarms (Table 4) were analyzed with a 2 × 2 × 2 ANOVA, with test item (O or P) and testing delay (short or long) as within-subjects factors and material (prose or poetry) as a between-subjects factor. The main effect of test item was significant [$F(1,41) = 46.50$, $MS_e = 0.037$, $p < .0001$], indicating that the participants distinguished targets from

lures. The interaction between test item, testing delay, and material was significant [$F(1,41) = 4.58$, $MS_e = 0.028$, $p < .05$]. For poetry, hits increased and false alarms decreased slightly over time, neither significantly. For prose, hits decreased significantly [$F(1,22) = 7.25$, $MS_e = 0.031$, $p < .05$], whereas false alarms remained about the same. There were no other significant results.

Discussion

Experiment 4 replicated with visual presentation the differences between prose and poetry in memory for surface details over time: O/P discrimination decreased for prose, but not for poetry. This outcome refutes the potential criticisms that differences in the speaker's articulation of words and phrases in prose and poetry caused the observed differences and that the previous results were specific to the poetry of A. E. Housman. It supports an interpretation in terms of structural differences between poetry and prose, since the visual presentation controlled for acoustic or emotional differences in the oral productions by a human speaker. The overall decrease in performance with visual versus auditory presentation might indicate an increased level of difficulty, suggesting the possibility that reading (at least in this rigidly paced sequential presentation) is more difficult than listening.

GENERAL DISCUSSION

Previous research had shown that recognition memory requiring fine discrimination of surface information decays over time for prose (Sachs, 1967) but improves for music (Dowling et al., 1995; Dowling et al., 2001). For prose stories, our study replicated the result that memory for surface details of a sentence fades rapidly, as indicated by the listener's or reader's ability to distinguish O from P test items sharing the same meaning but differing in surface details (Experiments 2–4). However, memory for surface details in poetry (as measured by O/P discrimination) did not decline across delays of up to 28 sec (Experiments 1, 3, and 4). The data pattern for poetry resembled that for data obtained for music, where discrimination between O items and similar lures remained strong after 30 sec and even improved in some conditions (Dowling et al., 2001). In the following, we propose to discuss the poetry results, first, in comparison with prose, which shares linguistic meaning and syntactic structures with poetry, and second, in comparison with music, which shares structural similarities involving temporal organization and rhythmicity.

Poetry Versus Prose

Two types of information have to be remembered in language, whether prose or poetry: formal aspects (syntax) and meaning (semantics). For prose, Sachs (1967) and Gernsbacher (1985, 1990) suggested that surface information is retrieved only as long as necessary to understand the sentence, whereas meaning remains for a longer time. Meaning is the most important aspect, and formal surface structures help to communicate the gist, without having a particular role (even if they might become important in some contemporary writing styles). In poetry,
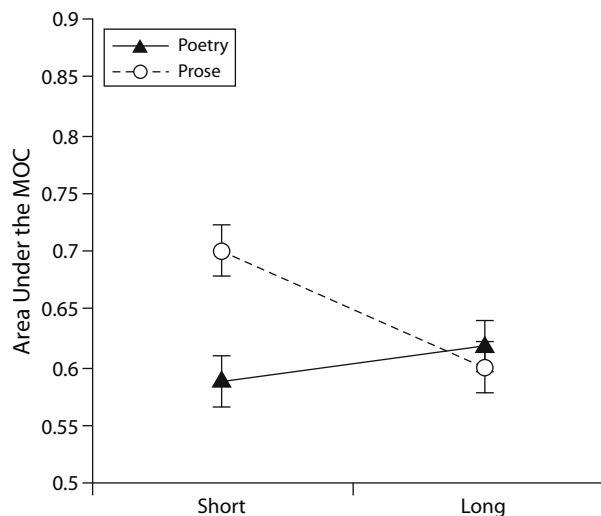

Figure 4. Area under the memory-operating characteristic (MOC) curve for O/P comparisons (comparisons of old verbatim items and paraphrased lures) at short and long delays in Experiment 4 for visually presented prose and poetry material. Short and long delays corresponded to about 9 and 25 sec, respectively. Error bars indicate between-subjects standard errors.

the importance of formal organization and sound patterns rivals the importance of linguistic meaning. The perceiver needs to grasp the specific surface realization—an intrinsic part of the poem—to appreciate it. This difference in the importance of surface structure might bring about the differences in memory for surface details between prose and poetry. The surface structures in question might be linked to the rhymes, the rhythmic organization, or their combination. Our study does not provide the means for disentangling these various influences, which will require future research. However, rhythm and rhyme have direct parallels in music (see below) and have been shown to affect verbal memory. For example, rhyme has been shown to lead to a decrease in recall for isolated words but to enhance recall for words in the context of sentences (Copeland & Radvansky, 2001). Our poetry data extend the influence of context on rhyming word memory (in recall) to the advantageous influence of rhyming words on sentence memory (in recognition).

Memory differences due to stimulus material have been reported for recall: Hypermnesia (recall improvement from an earlier to a later testing trial) is more pronounced for poems than for syllables or words (Ballard, 1913; Boreas, 1930, as reported by Erdelyi, 1996). We find similar results despite differences in method (recall vs. recognition) and delays (hours, days, months, or even years; see Erdelyi, 1996, for a review). Improvement and lack of decline of memory might be linked to the creation of richer and more precise memory traces, thanks to the stimulation of the intervening material (Tulving, 1983, 1984, 1986). For poetry, one feature of the material that would help to reinforce the stored features might be the rhythmic structures and their associated rhymes, which might provide global organization and useful cues (Rubin, 1995). Prose texts lack such memory aids.

Rhythm in Music and Language

Both poetry and music are strongly organized and structured over time. In poems, patterns of syllabic repetition and their associated stress (i.e., rhyme and rhythm) define structures that parallel such musical features as rhythm, meter, tonality, and tonal groupings (Lerdahl, 2003; Patel, 2003; Winn, 1981). Rhythmic structures in language (grouping of words into rhythmic chunks or periodicity due to stress) are reinforced in poetry, in comparison with prose, bringing poetry even closer to the rhythmic structures of music. Parallels in the rhythmic organization of poetry and music have led to the use of one system to describe the other: On the one hand, rhythmic structures in music have been described with the terminology for accents in poetry (see Clarke, 1989, for a review); on the other hand, Lerdahl proposed viewing poetry in terms of music. Analyzing a poem in terms of syntactic and prosodic structural stress, he drew a parallel between linguistic stress and perceived accents in music. More specifically, Lerdahl emphasized that in poetry and music, metrical structures consist of hierarchically related periodicities inferred by the perceiver. He further pointed out that the "mental representation of the sounds of metrical, rhymed poetry and of music . . . shares a good deal more organization than has usually been supposed" (p. 351). Parallels in temporal structures in music and poetry suggest the possibility that these features might be involved in the maintenance of memory traces over time, leading to the lack of decline for memory in both poetry and music.

Levels of Representation

If the argument is correct that temporal, especially rhythmic, organization contributes to the relatively long retention of the surface details of poetry, in comparison with prose, Kintsch et al.'s (1990) model of representation levels in discourse processing could be adapted to the present case. The model proposes abstract analyses of a text at surface (syntactic), propositional, and situational levels, explaining memory patterns over time. The relationships among linguistic elements (usually words) in the surface structure are described by a syntactic model. At the next level, a propositional textbase is constructed, capturing the overlapping arguments of the underlying propositions. Finally, there is the structure of the elements of the script describing the situation dealt with in the text (e.g., going to the movies). Each of these descriptive levels is characterized by a structural diagram representing the interconnections among the elements portrayed in the text.[1] The structural descriptions provide for deriving distances among the elements, and the distances are weighted depending on the degree of connectivity from element to element. A coherence matrix summarizes the strength of interconnectedness of the various elements across the text and across levels of analysis. The weights fall off rapidly with distance at the surface level, and less so in the propositional model. In the situational model, stronger weights are assigned to more remote links, capturing the intuition that scripts represent stable and tightly interconnected long-term memory structures. Memory performance can be predicted via the simulation of memory trace strength on the basis of the interconnections at surface, propositional, and situational levels.

In addition to the syntactic, propositional, and situational levels of description, we propose that the rhythmic patterns of poetry (and music) could be given a structural description along the lines suggested by Lerdahl (2003): a rhythmic/temporal structure. This structural description could be closely linked to the surface/syntactic description in the discourse-processing model. Since remote connections over relatively long structural distances are typical for the rhythmic organization of poetry, it might be the case that, like the connections at the situational level for prose, the weightings for elements in the rhythmic/temporal structure would extend over relatively broad distances in the model—broader, at least, than those assigned for the surface/syntactic structure alone. In this way, the resistance to forgetting of the surface details of poetry could be seen as part of the overall pattern of encoding and remembering text.

Influence of Closure on Memory

An additional feature of the similarity in temporal structure between poetry and music is the use of closure devices, such as rhyme and tonality. In poetry, recurring rhymes contribute to the structuring of the flow of the phrases, providing closure in the form of a predictable point of stability at the end of a phrase—a structure missing in prose. In music, establishing a tonality provides for the creation of local harmonic structures (cadences) that represent varying degrees of closure, structuring the musical flow and inducing patterns of tension and relaxation (Lerdahl & Jackendoff, 1983).

Gernsbacher (1985) has proposed a hypothesis based on processing shifts in relation to points of closure to explain the loss of surface information in stories over time. When processing stories, perceivers are building substructures to integrate and organize incoming information. As long as an episode (and its substructure) has not come to a close, the fine details of potential target items are held in working memory. At a point of closure, the substructure is closed, a new substructure is initiated, and surface information of the previous substructure is discarded. In poetry (and also in music), substructures might not be as clearly or as often closed as in prose. This might be the case because listeners need more contextual information to create a substructure or because rhythm and rhyme help make connections across potential closure boundaries. Both aspects would delay processing shifts and increase the time that detailed surface characteristics are available in memory. For most prose texts, surface information is the "most rapidly changing entity" (Gernsbacher, 1985, p. 351), whereas thematic content (at Kintsch's propositional and situational levels) is rather constant and is developed in memory even across processing shifts. Surface structure might be more stable in poetry and music than in prose because of rhythm and rhyme in poetry and coherent patterns of meter, rhythm, and melodic figures in music. Because of this stability, surface information is no longer the most rapidly changing entity and, hence, does not get lost with shifts. Finally, Gernsbacher has discussed the

possibility that surface information might be enhanced in memory, rather than suppressed over processing shifts, when the overall structure is more difficult to build (as for sentences with abstract vs. concrete content or with unusual syntax, as is often the case in poetry). This would suggest that structures of poetry and music require more complex processing than do those of prose—at least for the materials used in our experiments.

Feature Binding

For memory performance for poetry, false alarms declined over time for similar (P) lures, which was accompanied by decreasing hit rates (most strongly in Experiment 3). For music, a similar pattern underlies the observed improvement in the discrimination of details, with a decline of false alarm rates for similar (S) lures and with hit rates either being stable or slightly decreasing. These findings suggest that similar processes underlie the patterns of results for poetry and music. Decreased false alarm rates suggest some "sharpening" of the memory trace, accompanied by relatively strong hit rates for original items that are kept active and precise in memory.

Recent studies of recognition memory for words can shed some light on the processing steps leading to false alarms and their decrease over time (Jones & Jacoby, 2001; Kroll, Knight, Metcalfe, Wolf, & Tulving, 1996). For the avoidance of false alarms, Kroll et al. assigned an important role to the *binding* of stimulus features in early stages of trace consolidation; for them, correct encoding requires the binding of features into a coherent whole. For our study, considering *memory illusions* (P false alarms) in terms of feature binding suggests a different weighting of the binding process in poetry versus prose. Structures in poetry (and music) render salient some features (such as rhythm and rhyme) that are relevant for the target/lure distinction, and a long, continuous context helps bind the relevant features to the target, leading to a decline of P false alarms over time. These contextual-binding mechanisms of surface features will not operate for prose, since the features that lead to false alarms are largely irrelevant to the thematic sense and their integration is not supported by the ongoing context. This interpretation thus focuses on encoding and emphasizes the importance of the to-be-encoded features and their binding into a whole.

Conclusion

In this study, we investigated the time course of short-term memory for the surface details of verbal material. The similarity of the results for poetry and music provides us with some suggestions concerning the features influencing memory performance. Poetry and music share structural features involving rhythm (with underlying meter) and temporal structures creating closure. The parallel in memory performance over time between poetry and music (in contrast to prose) suggests that rhythmic structures and temporal organization are relevant features operating across material type (verbal vs. nonverbal). Notably, these features help to keep memory traces precise and may even increase their strength over time. This observation provides the basis for generating hypotheses for future research elaborating the way and the extent to which these features can aid listeners in retaining detailed information in memory. It is interesting to take note of the power of memory in oral traditions (i.e., stable memory for songs, stories, and poems for centuries, without the use of writing) and, especially, the role played by rhythm and music in that stability. Rubin (1995) called attention to the ways that oral traditions make use of these devices through materials that "are poetic, using rhyme, alliteration, assonance or some repetition of sound pattern, are rhythmic, are sung" (p. 8). Rubin discusses how patterns of sound can act as "constraints that cue memories and restrict choices" in recall (p. 65). He further states that "in most current psychological theories, words are kept in memory in an abstract form that loses all information about the sound pattern. But this shortcoming is remedied by the poetic devices, which constrain the sound pattern and thus help recover the original words" (p. 12). These observations lead us to summarize the present study as follows. In previous research and our experiments on prose, memory decay of detailed surface information over time has been observed. Our experiments on poetry showed this memory performance being maintained over time. The effectiveness of recall and reproduction in oral traditions leads us to the hypothesis that memory maintenance for fine surface details may even be enhanced when poetry and music are combined together in songs.

REFERENCES

BADDELEY, A. [D.] (1986). *Working memory*. Oxford: Oxford University Press, Clarendon Press.

BADDELEY, A. [D.] (1990). *Human memory: Theory and practice*. Boston: Allyn & Bacon.

BALLARD, P. B. (1913). Obliviscence and reminiscence. *British Journal of Psychology Monograph Supplements*, **1**, 1-82.

CLARKE, E. F. (1989). Considérations sur le langage et la musique. In S. McAdams & I. Deliege (Eds.), *La musique et les sciences cognitives* (pp. 23-44). Liège: Mardaga.

COPELAND, D. E., & RADVANSKY, G. A. (2001). Phonological similarity in working memory. *Memory & Cognition*, **29**, 774-776.

DOWLING, W. J., KWAK, S., & ANDREWS, M. W. (1995). The time course of recognition of novel melodies. *Perception & Psychophysics*, **57**, 136-149.

DOWLING, W. J., & TILLMANN, B. (2007). *Memory improvement while hearing music: Effects of structural continuity on feature binding*. Manuscript submitted for publication.

DOWLING, W. J., TILLMANN, B., & AYERS, D. F. (2001). Memory and the experience of hearing music. *Music Perception*, **19**, 249-276.

EBBINGHAUS, H. (1885). *Über das Gedächtnis*. Leipzig: Dunker.

ERDELYI, M. H. (1996). *The recovery of unconscious memories: Hypermnesia and reminiscence*. Chicago: University of Chicago Press.

FLETCHER, C. R. (1992). Assessing recognition memory for surface forms in discourse: A methodological note. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **18**, 199-203.

FLETCHER, C. R. (1994). Levels of representation in memory for discourse. In M. A. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 589-607). San Diego: Academic Press.

FLETCHER, C. R., & CHRYSLER, S. T. (1990). Surface forms, textbases and situation models: Recognition memory for three types of textual information. *Discourse Processes*, **13**, 175-190.

GERNSBACHER, M. A. (1985). Surface information loss in comprehension. *Cognitive Psychology*, **17**, 324-363.

GERNSBACHER, M. A. (1990). *Language comprehension as structure building*. Hillsdale, NJ: Erlbaum.

GLANZER, M., DORFMAN, D., & KAPLAN, B. (1981). Short-term storage in processing text. *Journal of Verbal Learning & Verbal Behavior*, **20**, 656-670.

HOUSMAN, A. E. (1963). *Collected poems*. New York: Holt.

JONES, T. C., & JACOBY, L. L. (2001). Feature and conjunction errors in recognition memory: Evidence for dual-process theory. *Journal of Memory & Language*, **45**, 82-102.

KINTSCH, W. (1994). The psychology of discourse processing. In M. A. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 721-739). San Diego: Academic Press.

KINTSCH, W., WELSCH, D., SCHMALHOFER, F., & ZIMNY, S. (1990). Sentence memory: A theoretical analysis. *Journal of Memory & Language*, **29**, 133-159.

KITAGAWA, H. (TRANS.) (1977). *Tale of the Heike* (2 vols.). New York: Columbia University Press.

KROLL, N. E., KNIGHT, R. T., METCALFE, J., WOLF, E. S., & TULVING, E. (1996). Cohesion failure as a source of memory illusions. *Journal of Memory & Language*, **35**, 176-196.

LERDAHL, F. (2003). The sounds of poetry viewed as music. In I. Peretz & R. J. Zatorre (Eds.), *The cognitive neuroscience of music* (pp. 413-429). Oxford: Oxford University Press.

LERDAHL, F., & JACKENDOFF, R. (1983). *A generative theory of tonal music*. Cambridge, MA: MIT Press.

MEHLER, J., & CAREY, P. (1967). Role of surface and base structure in the perception of sentences. *Journal of Verbal Learning & Verbal Behavior*, **6**, 335-338.

PATEL, A. (2003). Rhythm in language and music: Parallels and differences. In G. Avanzini, C. Faienza, D. Minciacchi, L. Lopez, & M. Majno (Eds.), *The neurosciences and music* (Annals of the New York Academy of Sciences, Vol. 999, pp. 140-143). New York: New York Academy of Sciences.

RUBIN, D. C. (1995). *Memory in oral traditions: The cognitive psychology of epic, ballad, and counting-out rhymes*. New York: Oxford University Press.

SACHS, J. S. (1967). Recognition memory for syntactic and semantic aspects of connected discourse. *Perception & Psychophysics*, **2**, 437-442.

SCHMALHOFER, F., & GLAVANOV, D. (1986). Three components of understanding a programmer's manual: Verbatim, propositional, and situational representations. *Journal of Memory & Language*, **25**, 279-294.

SWETS, J. A. (1973). The relative operating characteristic in psychology. *Science*, **182**, 990-1000.

TULVING, E. (1983). *Elements of episodic memory*. Oxford: Oxford University Press, Clarendon Press.

TULVING, E. (1984). Précis of Elements of episodic memory. *Behavioral & Brain Sciences*, **7**, 223-268.

TULVING, E. (1986). Episodic and semantic memory: Where should we go from here? *Behavioral & Brain Sciences*, **9**, 573-577.

VAN DIJK, T. A., & KINTSCH, W. (1983). *Strategies of discourse comprehension*. New York: Academic Press.

WAUGH, N., & NORMAN, D. A. (1965). Primary memory. *Psychological Review*, **72**, 89-104.

WILLIAMS, O. (ED.) (1947). *A little treasury of great poetry, English and American, from Chaucer to the present day*. New York: Scribner's.

WINN, J. A. (1981). *Unsuspected eloquence: A history of the relations between poetry and music*. New Haven, CT: Yale University Press.

NOTE

1. For example, "Nick" as the structural element representing the subject of the sentence "Nick decided to go to the movies," NICK as an element in the proposition DECIDE (NICK, (GO (NICK, MOVIES))), and "Nick" as the element representing the protagonist in a movie-going script all refer to the same element in the listener's or reader's memory representation, but that element is linked to the other elements in the text (the decision, the movies, etc.) in different ways, depending on whether we are considering the surface, the propositional, or the situational structure in the analysis.