

## Physical interaction and association by contiguity in memory for the words and melodies of songs

ROBERT G. CROWDER and MARY LOUISE SERAFINE  
Yale University, New Haven, Connecticut

and

BRUNO REPP  
Haskins Laboratories, New Haven, Connecticut

Three experiments were designed to investigate two explanations for the integration effect in memory for songs (Serafine, Crowder, & Repp, 1984; Serafine, Davidson, Crowder, & Repp, 1986). The integration effect is the finding that recognition of the melody (or text) of a song is better in the presence of the text (or melody) with which it had been heard originally than in the presence of a different text (or melody). One explanation for this finding is the physical interaction hypothesis, which holds that one component of a song exerts subtle but memorable physical changes on the other component, making the latter different from what it would be with a different companion. In Experiments 1 and 2, we investigated the influence that words could exert on the subtle musical character of a melody. A second explanation for the integration effect is the association-by-contiguity hypothesis, which holds that any two events experienced in close temporal proximity may become connected in memory such that each acts as a recall cue for the other. In Experiment 3, we investigated the degree to which simultaneous presentations of spoken text with a hummed melody would induce an association between the two components. The results gave encouragement for both explanations and are discussed in terms of the distinction between encoding specificity and independent associative bonding.

The concept of a song denotes a single entity with two constituents—words and melody. Questions about the cognitive representations of these theoretically separable, but subjectively intimate, elements of a song motivated our earlier investigations (Serafine, Crowder, & Repp, 1984; Serafine, Davidson, Crowder, & Repp, 1986). We suggested that a song could, logically, be represented in memory in three ways: (1) *independent storage* of components (the separate entities perceived and stored so that memory for one is uninfluenced by the other), (2) *wholistic storage* (the two components so thoroughly connected in perception and memory that one can be remembered only in the presence of the other), and (3) *integrated storage* (the two components related in memory such that one component is better recognized in the presence of the other than in its absence). The wholistic hypothesis can be dismissed out of hand, for people can often recognize the melodies of familiar songs when they are performed on solo instruments or with unfamiliar verses. What this informal observation leaves open, however, is whether the memory representation consists of independent or integrated components.

In our earlier studies, we reported evidence for what we called an *integration effect* in memory for the melody and text of songs. Using a recognition task, we found that melodies were better recognized when they were heard with the same words (as originally heard) than when they were heard with different words, even when the different words fit the melody and were equally familiar to the subject. Similarly, we found that the words of songs were better recognized in test songs containing the original melody than in those containing a different but equally familiar melody.

The procedure we employed was as follows. Subjects heard a serial presentation of up to 24 unfamiliar folk-song excerpts, each presented once. A recognition test followed immediately, in which the subjects were typically asked to indicate, for each excerpt, whether they had heard exactly that melody (or text) before, ignoring the current text (or melody). The test excerpts consisted of "old songs" (exactly as heard in the presentation) and various types of "new songs" (e.g., old melody with new words), including a type we termed "mismatch songs"—that is, an old melody with old words that had been paired with a different melody in the original presentation. The critical comparison was between melody recognition when old songs were tested and when mismatch songs were tested—that is, when the melody was paired with its original companion as opposed to a different, but equally familiar, one. This comparison, then, avoided the potentially biasing effect that completely new, unfamiliar words

This research was partially supported by NSF Grant GB 86-08344 to R. Crowder and by NICHD Grant HDO 1994 to Haskins Laboratories. We appreciate the assistance of William Flack in testing subjects and the comments of Shari Speer on an earlier version of this paper. Reprint requests may be sent to the first author at the Department of Psychology, Box 11A Yale Station, New Haven, CT 06520-1774.

might have had on recognition for a truly remembered melody. What we have termed the integration effect is the finding that both melody and text recognition were better in the case of old songs than in mismatch songs. We concentrated on the facilitating effect of identical words upon recognition of melodies, because recognition of words was in some cases almost at ceiling, although exactly comparable operations would be possible with recognition testing of the words from old songs, mismatch songs, and so on.

The effect was robust. It was not eliminated by instructing the subjects, on their initial hearing, to focus their attention on the melody (and ignore the words), nor was it eliminated by having a different singer on the recognition test than had sung in the original presentation (Serafine et al., 1984). Moreover, the effect was not due to the potential confusion of hearing the melody with seemingly "wrong" words; the mismatched words did not make melody recognition suffer, relative to an appropriate baseline (hummed melodies), but the original words truly facilitated it (Serafine et al., 1986). Thus, the difference between the old-song condition and the mismatch condition cannot be attributed simply to a decrement caused by mismatching familiar words and melodies.

Nor was integration accounted for by a semantic connotation imposed on the melody by the meanings of the words (Serafine et al., 1986), because the effect was found even in songs employing nonsense syllables on presentation and test. A melody heard only once, then, was better recognized in the presence of its original nonsense text than with different but equally familiar nonsense. This observation seems inconsistent with a meaning-interaction hypothesis, which might otherwise have had considerable intuitive appeal.

In the present study, we continued our exploration of the cognitive representation for remembered songs. Two hypotheses, not necessarily incompatible with one another, were under test here: the *physical interaction* hypothesis and the *association-by-contiguity* hypothesis.<sup>1</sup> The first of these hypotheses asserts that the integration effect is caused by physical changes, resulting from different phonetic properties of different texts. When a song is sung, the words impose subtle effects upon the melody notes, slightly changing such acoustic properties as their onsets, durations, and offsets, and perhaps their timbres as well. For example, some words might impose a staccato articulation and others a legato phrasing. (Compare singing any single melody, preferably a slow one, to the words "pitter-patter ..." with the words "singing only ..." to appreciate this.) We have termed these effects *submelodic*, because they leave unaffected the pitches and durations as they would be notated on a printed score.

If this submelodic hypothesis is correct, then a melody sung with one particular text would in fact become a somewhat different melody than it was when sung with another text, although not different on a printed score. It would not be surprising, then, to find the melody to be better recognized with the same words than with changed words.

A similar argument could be made for texts if we were concentrating on memory for song lyrics: different melodies change the prosody of the speech, even though the printed version of the speech stays the same.

The association-by-contiguity hypothesis is the Aristotelian claim that two events occurring in close temporal contiguity (successively or simultaneously) tend to become connected in memory, although neither event was necessarily changed by virtue of having entered into this association. If this hypothesis were correct, then, in the limit, text and melody would be associated just as well whether they were experienced simultaneously but separately (e.g., words spoken and hummed melodies) or were given as a song.

In the present research, the first two experiments were designed to address the submelodic hypothesis (a special case of the physical interaction hypothesis in which words are held to affect musical properties of the melody) and the last experiment was designed to address the association hypothesis. All three experiments employed our usual general procedure: subjects heard folksong excerpts followed immediately by a melody-recognition test in which the test items contained controlled combinations of song components. All three experiments used variations of the musical materials and design described below.

## GENERAL METHOD

Our musical materials were based on 40 American folksongs (from Erdei, 1974) which, in earlier experiments, we found were virtually all unfamiliar to our subjects (see the appendix of Serafine et al., 1984, for the list of songs). Eighteen pairs of song excerpts were used, each pair selected so that melodies and texts were interchangeable, in the sense of having rhythmic compatibility. Figure 1 shows such a pair. Interchangeability of melodies and texts was crucial to the construction of test items in which a song could contain a different melody or text than that heard originally in the presentation. Thus, each text contained a stress pattern suitable for either of the two melodies, and both texts within a pair contained the same number of syllables. The exceptions were two pairs, in which one text was shorter by one syllable and required the common "slur" across two tones (see "sleep" in Figure 1, Melody B). Given interchangeable components, we could arrange for the three test conditions: (1) old melody/old words, correctly paired, called "old songs"; (2) old melody/old words, "mismatch" songs; and (3) new melody/old words.

In each of the present three experiments, the critical comparison was between old songs and mismatch songs, where the latter items allowed us to test recognition of one component in the presence of a different component, which had nevertheless been heard in the presentation and was equally familiar. The third condition provides a baseline for false alarms that might be elicited by recognition of the words per se.

The songs were sung in the alto range by the second author, recorded onto a master tape, and dubbed onto sets of experimental tapes with a 5-sec interval of silence between presentation items and a 10-sec response interval after each test item. A silent metronome set at one beat per second facilitated performance at an even tempo, and a piano tone (not heard by the subjects) ensured pitch accuracy at the start of each song.

Subjective tempos across the songs were not uniform, however, due to normal rhythmic and metric variations (e.g., "double time"). The presented songs ranged in total duration from about 4 sec to

Melody	Text	
A		
	a	When the train comes a-long, When the train comes a- long.
	b	Hush a- bye, don't you cry, go to sleep lit- tle babe.
B		
	b	Hush a- bye, don't you cry, go to sleep lit- tle babe.
	a	When the train comes a- long, When the train comes a-long.

Figure 1. Two melodies, each with the same two interchangeable texts.

about 10 sec, with a mean of 6.4 sec and a standard deviation of 2.01 sec. All songs used C as the tonic, although there were variations in mode (Dorian, major and minor) and in starting note. A few slight alterations were made in the original folk melodies or texts (e.g., "across" was changed to "cross") in order to ensure rhythmic interchangeability of materials.

The same general design was used in all three experiments. In the presentation and test sequences, the song pairs were always heard in the same order. On the presentation tapes, half the songs were melodies with their original folksong texts (Type Aa in Figure 1) and half used the borrowed, interchangeable text (Type Ab in Figure 1). This precaution assured us that any advantage of old songs over mismatched songs could not result from the inherently greater compatibility of the original words and melodies in the old songs. Each mismatch item on the test tapes required two songs in the presentation sequence (since the melody of one would be tested with the text of another). Whenever two such songs occurred in the presentation, they followed one another immediately on the tape. Natural sources of variation among these songs include length, nature of the melody, tempo, and subject matter of the text, to name only a few characteristics. These factors were completely controlled, however, by counterbalancing across different subject groups.

Experiment 1 was mainly an effort to replicate our earlier finding of an integration effect with nonsense text. The aim of Experiment 2 was to test the submelodic hypothesis by employing, on the presentation and test tapes, different texts with similar phonetic patterns. We derived phonetically similar texts by translating each text into two corresponding nonsense texts, in which vowels were left intact and consonants were changed to reasonably close phonetic neighbors (e.g., /p/ = /t/ = /k/, /b/ = /d/ = /g/; see below). The presentation consisted of songs with nonsense texts, and the test consisted of songs with different but phonetically similar nonsense texts. If the submelodic hypothesis is correct, the integration effect should still be obtained. That is, a melody should be better recognized when it appears with meaningless syllables that are phonetically similar to those with which it was originally heard than with texts whose phonetic derivatives are equally familiar but had been heard originally with a different melody.

## EXPERIMENT 1

In planning the experimental effort described above, we developed a new and improved set of phonetic similarity

rules to use in deriving nonsense texts from one another. These were different enough from those used previously by Serafine et al. (1986, Experiment 1) that we wanted to verify that the original integration effect (i.e., with no phonetic discrepancy between the texts in the original presentation and test) would occur with them. In addition to confirming the suitability of the new rules, this would allow a replication of one of the more important results from earlier in the series: We found, as stated above, that recognition was better in the old-song condition than in the mismatch condition even when the texts were nonsense syllables. We thus falsified expectations that might have been based on a semantic, or meaning-compatibility, hypothesis, holding that the meanings of the words somehow combine with the "meaning" of the melody. This idea is intuitive enough, especially when applied to art songs, for example, that we thought it wise to repeat the comparison.<sup>2</sup>

## Method

**Materials.** The design of Experiment 1 required 18 song pairs. Because two nonsense texts were required for each real text in Experiment 2 (reported below), it was necessary to employ rules for translating real words into two similar nonsense texts. The rules, as follows, were similar to, but more detailed than, those used in Experiment 1 of Serafine et al. (1986) for the generation of a single nonsense text. For example, the voiced/unvoiced distinction was preserved across transformations, and manner classes were generally maintained. These rules allowed for at least two nonsense derivations from the original words:

(1) Vowels remained the same, and the following vowel-liquid sequences were treated as intact: /er/ as in *Mary*, /ar/ as in *far*, /il/ as in *will*, /or/ as in *lore*, /ol/ as in *boy*, /aU/ as in *how*, /eIl/ as in *pail*, /al/ as in *doll*, /ol/ as in *awl*, and /anz/ as in *runs*.

(2) Consonants were interchanged according to the following list of phonetic similarities. For example, if /b/ occurred in a real word, the two corresponding nonsense words used /d/ and /g/, respectively,

/b/ = /d/ = /g/

/n/ = /m/ = /l/

/r/ = /l/ = /j/ or /w/; /w/ = /j/ = /r/ or /l/

/p/ = /t/ = /k/  
 /f/ = /θ/ = /h/ or /s/; /θ/ = /f/ = /ʃ/ or /s/; /h/ = /θ/ = /f/;  
 /s/ = /ʃ/ = /f/  
 /z/ = /v/ = /ð/  
 /tr/ = /kw/ = /pl/  
 /pr/ = /tw/ = /kl/  
 /kr/ = /tw/ = /pl/  
 /st/ = /sk/ = /sp/  
 /sl/ = /ʃw/ = /fr/  
 /skw/ = /str/ = /spl/  
 /bl/ = /dw/ = /gr/  
 /sp/ = /st/ = /sk/  
 terminal /n/ = terminal /m/ = terminal /ŋ/.

Special cases of translated vowel/consonant combination:

/ɜr/ = /ɜrm/ = /ɜrn/ (e.g., *girl* = *berm* = *dern*)  
 /ir/ = /il/ = /in/ (e.g., *here* = *seal* = *feen*)  
 /end/ = /emd/ = /emb/  
 /ʌn/ = /ʌm/ = /ʌl/  
 terminal /ɔŋ/ = /ɔn/ = /ɔm/ (e.g., *song* = *fawn* = *shawm*)  
 terminal /ŋ/ = /Im/ = /In/

(3) Interior /ɜr/ or /ɜn/ was treated as a vowel, but in terminal position it was interchanged as follows: /ɜr/ = /ɜn/ = /ɜl/ as in the second syllable of the words *anger*, *often*, and *able*.

(4) A terminal /s/ or /z/, when a plural marker, was retained (untranslated) if the resulting nonsense was too difficult to pronounce.

(5) The sounds /tʃ/ and /dʒ/ were omitted from all texts because three suitable phonetic correspondences do not exist. Thus, minor changes in some original texts were made (e.g., *Joe* was changed to *Moe*, *chase* was changed to *run*).

The following is an example of a translated text, written in the form (regular orthography) used by the singer:

Original: Cobbler, cobbler make my shoe.  
 Nonsense 1: Poggrel, poggrel nate nie foo.  
 Nonsense 2: Toddwen, toddwen lape lie thoo.

**Design.** Three sets of presentation and test tapes were administered to different groups of subjects. Across the three subject groups, each presentation item was tested in each of the three conditions: old song, mismatch, and new melody/old words. The presentation tapes consisted of 24 songs, and test sequences consisted of 18 items, six each of the three conditions.

**Procedure.** The subjects were instructed to listen to a presentation of folksong excerpts with nonsense texts, were told that their memory would be tested later, and, following the presentation list, were given a melody-recognition test in which they were to indicate whether they had "heard this exact melody before—that is, just the musical portion." They were not told what types of items to expect on the test except that the nonsense folksongs would be similar to those on the presentation.

**Subjects.** Fifteen Yale undergraduates with undetermined levels of musical training were equally divided among the three groups.

## Results and Discussion

Responses were translated into 6-point "oldness" ratings, where 1 represents very confident no (had not heard the melody) and 6 represents very confident yes (had heard the melody). The use of rating scales such as the present one is commonplace in recognition-memory situations (Murdock, 1974, chapter 2) and is known not to differ in measured sensitivity from yes-no techniques (Pollack & Decker, 1958), only in efficiency (because with six rating categories, subjects are applying five criteria at the same time).

Table 1  
Mean Oldness Ratings by Conditions and Experiments

	Test Condition		
	Old Song	Mismatch	New Melody/ Old Words
Experiment 1	4.85	3.63	2.59
Experiment 2	4.26	3.88	3.05
Experiment 3:			
Divided songs	4.56	4.04	3.33
True songs	4.35	3.96	3.25

Note—Mean square error terms for the main effect of the three conditions across the three experiments were, respectively, .452, .404, and .377.

The results are shown in the first row of Table 1. Two omnibus analyses of variances yielded statistically significant results: with subjects as the sampling variable [ $F(2,28) = 42.66, p < .001$ ] and with the 18 test items as the sampling variable [ $F(2,34) = 41.76, p < .001$ ]. Newman-Keuls tests revealed that melody recognition was significantly better in the old-song condition than in the mismatch condition, both across subjects ( $p < .01$ ) and across items ( $p < .01$ ).

Thus, the integration effect was obtained with these new nonsense materials, verifying that the presence of the original old text—even nonsense words—facilitates melody recognition over that obtained with a different but equally familiar text, in the mismatch songs. In addition to confirming our new stimuli as appropriate for the tasks, the results of Experiment 1 provide a welcome replication of one of our most important previous results: In Experiment 1 of Serafine et al. (1986), the means corresponding to the first row of Table 1 were 4.47, 3.76, and 2.60.

## EXPERIMENT 2

The aim of Experiment 2 was to test the submelodic hypothesis directly, the preconditions for this test having been realized in Experiment 1. We employed two different sets of phonetically derived nonsense texts, based on the same real words (which were never used in this experiment). The presentation consisted of folksong excerpts with nonsense texts. The test consisted of folksong excerpts whose texts were phonetically similar to those in the presentation but nevertheless were, in all cases, different nonsense. (As in Experiment 1, the phonetic derivative of an old song was called an "old song," etc., for the purposes of this experiment.) Test items were of three types: (1) "old songs" (old melody with nonsense words that were phonetically similar to the old nonsense text), (2) "mismatch songs" (old melody with nonsense words that were phonetically similar to an old nonsense text from a different song in the presentation), and (3) new melody/"old words" (new melody with nonsense words that were phonetically similar to an old nonsense text).

If the submelodic hypothesis is correct, then a melody should be better recognized when it is heard with non-

sense words that are phonetically related to the nonsense with which that melody was originally presented than when it is heard with nonsense that is not phonetically related to the original. In other words, melody recognition in "old songs" should exceed that in "mismatch songs."

### Method

**Materials.** The materials were those described in Experiment 1. Both sets of nonsense texts (phonetic derivatives of the original folk-song texts) were employed.

**Design.** The design was comparable to that of Experiment 1, except that instead of comprising old songs, mismatch songs, and new melodies with old words, the test items used the phonetically derived "old songs," "mismatch songs," and new melodies with "old words," where our quotation marks indicate that *exact* repetition of the verbal texts between presentation and test never occurred. As in Experiment 1, counterbalancing across subject groups was employed to control for natural variations in the songs. The presentation consisted of 24 items and the tests consisted of 18 items.

After 12 of the 30 subjects had been tested, an inadvertent error in the test tapes was detected. Two song pairs contained faulty material for the new melody/"old words" condition, although the other two conditions were correct. Thus, scores for those 12 subjects were based on four (instead of six) items in the new melody/"old words" condition.

**Procedure.** The procedure was analogous to that of Experiment 1. At test, the subjects were told that the texts of songs may sound similar to or different from those heard before, but they were to attend only to the melody and indicate recognition (yes or no) and a confidence rating on the answer sheet.

**Subjects.** Thirty college-age adult volunteers with undetermined levels of musical training were paid to participate and were equally divided among the three groups.

### Results and Discussion

As in Experiment 1, melody-recognition ratings had a possible range of 1-6. Means for the "old song," "mismatch," and new melody/"old words" conditions are shown in the second row of Table 1. With subjects as the sampling variable, the result of an analysis of variance was significant [ $F(2,58) = 28.18, p < .001$ ], and Newman-Keuls tests indicated that melody recognition under the "old song" condition was significantly better than that under the "mismatch" condition ( $p < .05$ ).

With items as the sampling variable, an analysis of variance was performed on the means generated by the 18 subjects who had completed all items in all conditions. Those means were 4.10, 3.71, and 3.04, respectively, for the "old song," "mismatch," and new melody/"old words" conditions. The main effect was significant [ $F(2,34) = 12.02, p < .001$ ], and a priori comparisons involving only the first two conditions revealed significance at the .02 level. (The results of post hoc tests were not significant for this item analysis, however.)

The results of the present experiment show that the integration effect is obtainable with phonetically similar, as well as identical, nonsense used at test. In our view, it is indisputable that words (whether meaningful or nonsense) exert variable effects on melody tones, as can be easily imagined with the "pitter-patter ..." and "sing-

ing only ..." thought experiment suggested above. Anyone who has glanced even casually at speech spectrograms knows how the unvoiced stops in the former phrase interrupt the vocal stream with periods of actual silence, unlike the latter phrase. The present results suggest that these submelodic changes are retained in memory and are at least partially responsible for the integration effect.

We think it is not an accident that the present experiment showed evidence favorable to the submelodic hypothesis, whereas some earlier, unpublished efforts with a similar experimental design did not.<sup>3</sup> The rules for deriving phonetically similar nonsense texts were more fastidious here than those used before. For example, in these new materials, we respected the voiced/voiceless distinction more consistently than under the old rules. Consonants with full stop closures were distributed equally in the original and derived versions, too. These distinctions are just the sort that would be expected to underlie a submelodic effect of words on music.

Other interpretations of the integration effect (e.g., those to be considered below) might also be consistent with the evidence adduced here for the submelodic hypothesis. In comparing Experiments 1 and 2 of the present series, we note a numerically smaller, and statistically weaker, integration effect in the latter experiment (with the derived nonsense words) than in the former (with the very same nonsense texts presented at learning and testing). This is as it should be, in any commonsense view, because no scheme for deriving "similar" phonetic texts could possibly be as faithful a reinstatement as complete identity. On the other hand, we should not exaggerate the triumph of the submelodic hypothesis: at most, we can claim that we have shown conclusively that some such factor was operating somehow in our integration experiments, not that it is a complete explanation of the effect.

### ON ASSOCIATIONS

Experiment 3 (below) was designed to address what we have referred to as the association-by-contiguity hypothesis. The term "association," by itself, may connote many things theoretically, such as rote learning, Pavlovian conditioning, or the antediluvian mists of precognitive psychology. However, the term is theoretically empty; it simply connotes an experimental fact, that Events A and B stand in a particular empirical relationship because of their history of co-occurrence. The challenge for theory is to rationalize the circumstances necessary for that association to be formed and expressed, and the nature of the bonding thereby achieved. Thus, our integration results undoubtedly illustrate some form of association. The submelodic mechanism, for which we adduced some support in Experiment 2, is not strictly an associative mechanism at all, but rather is an effect of one element upon the identity of the other—namely, that the occurrence of A with B changed the physical nature of B. We now ask whether the temporal contiguity of A and

B (words and melody, respectively) is a sufficient condition for their association when no possibility exists for an overt influence of one upon the physical integrity of the other.

In considering the theory of associations, we have relied upon the distinction in the respective psychologies of James Mill and John Stuart Mill between *mental compounding* and *mental chemistry* (see Boring, 1957, chapter 12). In the former case, two components retain their independent identities yet are connected to one another. In the latter case, the two components are themselves altered by each other's presence. Our concept of the association by contiguity of melody and text is like that of mental compounding: the melody and text are connected in memory, hence they act as recall cues for each other, yet each is stored with its independent integrity intact. By contrast, the submelodic hypothesis is consistent with a "chemical" form of bonding, holding that a melody and text change each other physically when sung together in a song. A more purely *mental chemistry* could be an associative process in which, by co-occurrence in the mind, the memory representation of each is changed, relative to what it would have been without a particular companion.

More recently, a similar distinction has been articulated by Horowitz and Manelis (1972), albeit with a linguistic orientation, for adjective-noun phrases. They refer to a distinction between *I-bonding* (where I stands for *individual* or *independent*) and *J-bonding* (where J stands for *joint*). The former, illustrated by the phrases *deep chair* or *dark wing*, take their meaning as a phrase from the meanings of the constituent words. The latter, illustrated by *high chair* or *right wing*, possesses idiomatic meaning that transcends the meanings of the two constituents. As Horowitz and Manelis remarked (p. 222), *I-bonding* owes allegiance to the British empiricist philosophers and *J-bonding* to the Gestalt tradition. Tulving's work on recognition failure in episodic memory (Tulving, 1983) illustrates the same properties as *J-bonding*, wherein an element of an association may be only poorly recognized but can be well recalled given the original associate as a cue. In many ways, we believe that these issues are raised in stark relief when the two constituents, such as words and melodies, are fundamentally different cognitive elements as opposed to when verbal associations are at stake.

### EXPERIMENT 3

In the present experiment, we investigated whether temporal contiguity is a sufficient condition for association. We assessed the degree to which a text could serve as the retrieval cue for a melody, when the two had initially been heard in closed temporal proximity (independent but simultaneously presented). The main question was whether the old-song condition would generate better melody recognition than would the mismatch condition. If a simple contiguity hypothesis is correct, we would expect

that independent melodies and texts, presented together, would become connected in memory, and hence could act as recall cues for one another. Finding integration with our distinct-but-contiguous presentation of words and melodies (see below) would be a failure to reject the mental compounding idea, and would thereby provide no comfort for the chemical analogy in association theory.

The presentation episodes consisted of normal spoken texts and hummed melodies heard simultaneously and binaurally (but not dichotically). We refer to these simultaneous pairings as "divided songs." The recognition tests were of two types: half the subjects heard only divided songs (as in presentation) and half heard true, sung songs. No instruction for the generation of song-like representations in presentation was given. So the question at hand was whether an association between contiguous components, if it occurred, would influence melody recognition only if the test stimuli were like those of the presentation, or whether that association's influence would also extend to the case of true songs.

The melody-recognition tests for both of the (between-subject) conditions with divided songs and true songs consisted of three within-subject conditions. As before, the critical comparison was that between old songs and mismatch songs, with the old words/new melody condition added as a baseline comparison.

### Method

**Materials.** Eighteen song pairs were used. A master tape, from which experimental tapes were dubbed, was prepared by the same alto singer, as follows. Hummed melodies were first recorded in succession, each preceded by exactly four evenly spaced taps, which were also recorded onto the tape. The resulting signal was then fed into a second tape recorder at the same time that spoken texts were recorded onto a second tape. The singer listened to the hummed melodies from the first tape over headphones, using the four taps to fix the onset of the hummed melody, and then spoke the text along with the melody, recording both onto the second tape. Texts were generally spoken in the rhythm of the melody and also began and terminated in synchrony with it. When experimental tapes were dubbed from the master, the four taps were omitted. The test tapes in the spoken-song condition were dubbed from the same master tape as the presentation tapes; those employing true songs were available from previous experiments using these materials.

**Design.** The design was exactly analogous to that of Experiments 1 and 2, except that two sets of test tapes were constructed, each administered to a different group of subjects, one set with divided songs and the other with true songs.

**Procedure.** The procedure was comparable to that of the earlier experiments. The subjects were told to expect the spoken texts of simple folksongs to be heard simultaneously with hummed melodies. At test, one group was told that items would be true, sung songs, and the other group was told that test items would be similar to presentation items. In all cases, of course, the instructions called for recognition based only on the melodies.

**Subjects.** Twenty-four adults with undetermined levels of musical training were equally divided between the two test groups.

### Results and Discussion

Melody-recognition ratings had a possible range of 1-6. Mean ratings for old songs, mismatch songs, and old words/new melody, respectively, are shown in the last

two rows of Table 1. Two mixed analyses of variance were performed with type of test (spoken vs. true songs) as a between-subjects variable and the three conditions (old song, mismatch, and new melody/old words) as a within-subjects variable. With subjects as the sampling variable, only the main effect of conditions was significant [ $F(2,44) = 21.68, p < .001$ ]; neither type of test nor the interaction was significant. The Newman-Keuls test indicated that combined old-song ratings for both groups (4.46) exceeded those for mismatch songs (4.00) ( $p < .05$ ). Similarly, with items as the sampling variable, only differences among the three conditions were significant [ $F(2,68) = 13.65, p < .001$ ]. The Newman-Keuls test again supported the difference between old-song and mismatch ratings ( $p < .05$ ).

Thus, by the reasoning explained above, true temporal contiguity of melody and text was a sufficient condition for observing the integration effect. Preliminary experiments, involving several arrangements of *successive* contiguity, had failed to yield integration, even when subjects had been instructed to try to imagine the words and music as a song. The most straightforward interpretation of this result is that, in Experiment 3, simultaneous presentation was favorable for the formation of associative links between constituents that had not lost their individual identity. Here, for the first time in this series, we may categorically rule out the submelodic hypothesis, because the pairing manipulation could not have had any substantial effect on the physical nature of each constituent.

The argument could be made, however, that our subjects all emerged from the presentation sequence with self-generated songs as memory representations. Hearing a hummed melody at the same time as one hears a rhythmically compatible stream of words might produce the experience of a song, whether the subject is deliberately trying to generate this song or not. Such a process would account for the integration effect among the subjects tested with real songs, because such test items would then match their memory residue from the original presentation. To account for integration among the subjects tested with divided songs, we need only observe that for these subjects, the conditions of acquisition and testing were precisely the same; this beneficial circumstance could have outweighed the disadvantage produced by the need for these subjects to generate song representations at test, as well as at acquisition.

An automatic process of generating song-like representations from simultaneous, compatible verbal and musical streams would not be unexpected from a consideration of speech processing. Because the segmental features of words are always overlaid upon suprasegmental features, specifically including variation in fundamental frequency, simultaneous variation in pitch might be automatically assigned to the prosodic aspect of speech, even when the listener "knows" the verbal and tonal messages are independent, as in listening to divided songs. Thus, even though Experiment 3 is most easily interpreted as

evidence for associations by sheer temporal contiguity, we cannot categorically reject the possibility of song generation. The burden of evidence now rests with those who wish to make such an assertion.

## GENERAL DISCUSSION

With regard to the integration effect in song, our experiments in this and in the two previous papers (Serafine et al., 1984; Serafine et al., 1986) have informed our thinking in a number of ways. First of all, and despite intuition to the contrary, the *meaning* of words seems to have a negligible role in the fact that melody and words become stored in an integral fashion. Here again, in Experiment 1, the result withstood nonsense materials devoid of conventional meaning.

Second, we have provided statistically reliable support for the submelodic hypothesis, which suggests that particular words can change the musical line sufficiently to influence later recognition of the melodies. People are understandably slow to realize that "baa, baa, black sheep" and "twinkle, twinkle, little star" are words to almost exactly the same tune—they are not, musically, quite the same tune, by virtue of the words to which each has been set, even though the tune might be identical on a written page. The melody as perceived and as remembered may be considerably richer than the abstraction represented by musical notation! (See comments by Bengtssen, 1987, on a similar consideration in rhythm, and by O'Connell, 1988, on the "written language bias" in linguistics.)

Finally, we have uncovered a number of factors that govern the size of the integration effect, some statistically reliable on their own and others not. Among the factors for which evidence exists, we must include first temporal contiguity, as shown in Experiment 3. Barring the unknown contribution of automatic fusion of text and melody in songs, hearing words and a melody at the same time appears to affect their joint storage in the manner of paired associates. But we should not completely discard factors uncovered by earlier experiments in this series as potential contributors to the effect. Even though they were not reliable on their own, they did measurably influence the size of the effect. Among these factors, we count (1) instructions to attend only to melodies rather than to the whole songs at presentation, and (2) acoustic nonidentity of presentation and test materials (i.e., different singers; Serafine et al., 1984). Elsewhere (Serafine et al., 1986), we found that (3) melodies in the presence of the *wrong* words did indeed have some distracting effect on melody recognition, beyond the facilitation that the correct words had. We have ruled out these three factors, individually, as necessary conditions for the integration effect, but we cannot pretend they are not correlated with it, collectively.

In these experiments, we did not try to evaluate the musical background of the subjects. In Experiment 1 of Serafine et al. (1984), such a comparison was possible, with negligible results (the main comparisons among ex-

perimental conditions were unaffected). In principle, we would expect that a developmental tendency to move from integral to separable perception of multidimensional stimuli would prevail (Pomerantz & Lockhead, 1990). Therefore, we would predict stronger integration for younger or musically inexperienced subjects than for older, more musically experienced ones, in an experiment designed to include a wide range of musical and chronological development. This might prove an interesting direction in which to take this research program in the future.

Putting all these factors together, we believe we know well how to arrange conditions in order to maximize, or minimize, the integration of words and melodies in recognition of songs. This laboratory control is arguably a form of explanation.

#### REFERENCES

- BENGTSSON, I. (1987). Notation, motion, and perception: Some aspects of musical rhythm. In A. Gabrielson (Ed.), *Action and perception in rhythm and music* (pp. 69-80). Stockholm: Royal Swedish Academy of Music.
- BORING, E. G. (1957). *A history of experimental psychology*. New York: Appleton-Century-Crofts.
- CROWDER, R. G., SERAFINE, M. L., & REPP, B. R. (1990). Physical interaction and association by contiguity in memory for the words and melodies of songs. *Status report on speech research* (No. 99/100). New Haven, CT: Haskins Laboratories.
- ERDEI, P. (1974). *American folksongs to read, sing, and play*. New York: Boosey and Hawkes.
- HOROWITZ, L. M., & MANELIS, L. (1972). Toward a theory of re-integrative memory: Adjective-noun phrases. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 6., pp. 193-224). New York: Academic Press.
- MURDOCK, B. B., JR. (1974). *Human memory: Theory and data*. Hillsdale, NJ: Erlbaum.
- O'CONNELL, D. C. (1988). *Critical essays in language and philosophy*. New York: Springer-Verlag.
- POLLACK, I., & DECKER, L. R. (1958). Confidence ratings, message reception, and the receiver operating characteristic. *Journal of the Acoustic Society of America*, 30, 288-292.
- POMERANTZ, J. R., & LOCKHEAD, G. (1990). *The perception of structure*. Hillsdale, NJ: Erlbaum.
- SERAFINE, M. L., CROWDER, R. G., & REPP, B. H. (1984). Integration of melody and text in memory for song. *Cognition*, 16, 285-303.
- SERAFINE, M. L., DAVIDSON, J., CROWDER, R. G., & REPP, B. H. (1986). On the nature of melody-text integration in memory for songs. *Journal of Memory & Language*, 25, 123-135.
- TULVING, E. (1983). *Elements of episodic memory*. New York: Oxford University Press.

#### NOTES

1. We are well aware that the dictionary meaning of the word "contiguity" stipulates that the events in question be juxtaposed, or adjacent, in time, but not overlapping or coterminous. This departs from usage of the term within psychology, where successive and simultaneous arrangements are both considered contiguous. In this paper, we remain with this latter usage even though the former might be more justifiable to some pedants.
2. In general, we deplore the reluctance of psychologists to perform straight replications of important results.
3. These earlier experiments are included in Crowder, Serafine, and Repp (1990) but were deleted from an earlier draft of the present article.

(Manuscript received February 24, 1989;  
revision accepted for publication January 11, 1990.)