Perception & Psychophysics
1995, 57 (2), 136–149

The time course of recognition of novel melodies

W. JAY DOWLING, SEYEUL KWAK, and MELINDA W. ANDREWS
University of Texas at Dallas, Richardson, Texas

Seven experiments explored the time course of recognition of brief novel melodies. In a continuous-running-memory task, subjects recognized melodic transpositions following delays up to 2.0 min. The delays were either empty or filled with other melodies. Test items included exact transpositions (T), same-contour lures (SC) with altered pitch intervals, and different-contour lures (DC). DCs differed from Ts in the pattern of ups and downs of pitch. With this design, we assessed subjects' discrimination of detailed changes in pitch intervals (T/SC discrimination) as well as their discrimination of contour changes (T/DC). We used both artificial and "real" melodies. Artificial melodies differed in conformity to a musical key, being tonal or atonal. After empty delays, T/DC discrimination was superior to T/SC discrimination. Surprisingly, after filled delays, T/SC discrimination was superior to T/DC. When only filled delays were tested, T/SC discrimination did not decline over the longest delays. T/DC performance declined more than did T/SC performance across both empty and filled delays. Tonality was an important factor only for T/SC discrimination after filled delays. T/DC performance was better with rhythmically intact folk melodies than with artificial isochronous melodies. Although T/SC performance improved over filled delays, it did not overtake T/DC performance. These results suggest that (1) contour and pitch-interval information make different contributions to recognition, with contour dominating performance after brief empty delays and pitch intervals dominating after longer filled delays; (2) a coherent tonality facilitates the encoding of pitch-interval patterns of melodies; and (3) the rich melodic–rhythmic contours of real melodies facilitate T/DC discrimination. These results are discussed in terms of automatic and controlled processing of melodic information.

When a melody is transposed in pitch, it remains the same melody as long as its intervals on a logarithmic frequency scale are preserved. This pattern constancy across a change in physical elements was often used as an example by the Gestaltists (e.g., Koffka, 1935). Adults and children without specific musical training readily agree that an exact transposition of a familiar melody is still the same melody. They find it easy to recognize a familiar melody presented at an arbitrary pitch level when its pitch intervals are preserved, and easy to detect alterations in those intervals (Dowling & Fujitani, 1971; Trehub, Morongiello, & Thorpe, 1985). In fact, musically untrained adults find it relatively difficult to identify a familiar melody when its intervals have been altered, even when its contour—the pattern of ups and downs—has been preserved (Dowling & Fujitani, 1971). In general, when we recognize a melody in everyday life, it is exact transpositions to arbitrary pitch levels that we recognize. Therefore, we take the recognition of a transposed melody as the basic standard for successful melody recognition.

Completion of this work was facilitated by Grant 27131-541 from the National Institute for Aging. We thank Hervé Abdi, James C. Bartlett, Robert G. Crowder, Andrea Halpern, and three anonymous reviewers for valuable suggestions. Requests for reprints should be addressed to W. J. Dowling, Program in Cognitive Science, University of Texas at Dallas, Richardson, TX 75083-0688 (e-mail: jdowling@utdallas.edu).

Since a transposition of a melody is one step further removed from the original than would be an exact repetition at the same pitch level, the repetition may seem to be a better candidate for the standard of melody recognition. There are two reasons why we opt for the transposition as a standard. First, the listener can detect alterations in a repetition by detecting alterations in its individual pitches, without taking account of the melodic pattern. Thus, if we want to test *melody* recognition, we need to use transpositions. Second, in everyday life, familiar melodies in long-term memory are readily recognized even though seldom presented at a constant pitch level from one occurrence to the next. (The NBC chime pattern is a rare exception, having been always presented with the same pitches; see Attneave & Olson, 1971.) What is important to melody recognition is the pattern of pitches, not the identities of the pitches themselves.

We can assess the relative importance to transposition recognition of various melodic features by removing features one by one and assessing the impact on recognition. To separate the contributions of melodic features such as contour and interval pattern to melody recognition, we use contrasts such as those shown in Figure 1. A melody is a pattern of pitches in time, and, at a basic level, the pitches themselves constitute a set of features.[1] However, as we have seen, a melody may be transposed to a new pitch level and remain the same melody even though all of its individual pitches have been changed. The melody T in Figure 1 is a transposition of the initial

Figure 1. Stimulus patterns typical of the experiments: A, a novel tonal melody; T, an exact transposition of melody A; SC, a same-contour lure with altered pitches indicated by bracket; DC, a different-contour lure.

melody A, preserving both contour and exact interval pattern of the original. SC in Figure 1 is a same-contour transformation of A in which some intervals have been changed. DC is a totally different melody with a different contour.

Listeners' ability to notice changes in the intervallic detail of familiar melodies contrasts sharply with their inability to notice such changes in novel melodies they have just heard (Dowling, 1978; Dowling & Fujitani, 1971). Immediate recognition of novel melodies is dominated by contour. Immediately following presentation of the target, SC lures tend to be confused with transpositions (Ts), whereas changes in contour are easily noticed, leading DC lures to be rejected.

It is tempting to attribute this qualitative difference in memory for familiar versus novel melodies to the overlearning of familiar melodies, which should produce a highly accurate representation of their intervallic detail. However, when novel melodies, presented just once, are tested after a delay filled with distracting material, contour declines in importance (compared with the immediate test), and the exact pitch pattern is more likely to be required for recognition (DeWitt & Crowder, 1986; Dowling & Bartlett, 1981; Edworthy, 1985). For example, Dowling and Bartlett (1981, Experiments 1 and 2) presented phrases of Beethoven quartets (unfamiliar to their high-school student listeners) and tested them more than 5 min later. Test stimuli were phrases from the same quartets whose relationships closely paralleled those of melodies A, SC, and DC in Figure 1. Listeners distinguished As from both SCs and DCs with about 75% accuracy (where chance was 50%). Thus, even with novel melodies heard just once, contour is much less important when tested after a filled delay than when tested immediately, and, following the delay, listeners are better able to distinguish between targets and highly similar SC lures than they are on immediate test. Even when listeners in Dowling and Bartlett's study were instructed to respond on the basis of contour, saying "yes" to both targets and SCs and "no" to DCs, they failed to follow the instructions, continuing to distinguish targets from SCs about as well as from DCs. This inability to follow instructions to override T/SC discrimination suggests the presence of an automatic process not subject to cognitive control. The results show that a shift in the relative importance of contour and exact intervals in melody recognition occurs across a filled delay even without the opportunity for listeners to overlearn the interval pattern of the target.

In a study that leads directly to the present experiments, Dowling (1991b) obtained results qualitatively similar to those of Dowling and Bartlett (1981) with tonal (but not with atonal) novel melodies, while controlling the length of delays more precisely. Dowling tested listeners' recognition of transposed melodies after empty delays of 11 sec and filled delays of 39 sec. Figure 2 shows the results. T/DC discrimination was superior to T/SC discrimination following the brief empty delay; however, T/SC discrimination improved for tonal (but not atonal) melodies following the longer filled delay.

The improvement in T/SC discrimination across a filled delay is puzzling, but it is highly plausible that it would be more likely to occur with tonal (vs. atonal) melodies. Tonal melodies should be much more easily and accurately encoded than atonal melodies, since they conform to the invariants of a diatonic scale pattern highly familiar to the listeners. Whether we believe (with Dowling, 1991b) that T/SC discrimination improves because of dissipation of confusions experienced

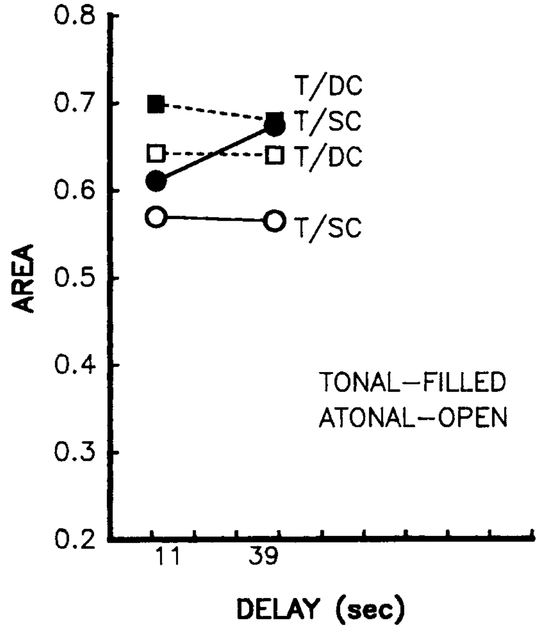

Figure 2. Areas under the MOC for the strongly tonal (filled symbols) and atonal (open symbols) conditions of Dowling (1991b) for melodies tested after an empty delay of 11 sec and a filled delay of 39 sec. Circles indicate discrimination between Ts and SC lures, and squares indicate discrimination between Ts and DC lures.

on immediate test as a result of similarities of contour and key between Ts and SCs or we believe that the improvement in T/SC discrimination is the result of covert rehearsal and strengthening of precise interval encoding, tonal melodies will have the advantage.

To decide theoretical issues such as the one just posed, we need more information. The present experiments were designed to address a series of questions arising from consideration of the above results.

(1) What would happen with the relative strength of T/SC versus T/DC discrimination with delays longer than 39 sec? Their curves are headed in opposite directions, but would they ever cross? If they did cross, it would pose the puzzle of how T/SC discrimination could ever be better than T/DC discrimination. How could the finer grained discrimination produce superior performance to the coarser, especially when the information required to carry out the coarser discrimination is logically implied in the finer discrimination? Experiment 1 extended Dowling (1991b) by increasing time delays to 1.5 min, and Experiment 6 increased them to 2.0 min.

(2) Both Dowling (1991b) and the present Experiment 1 confound delay and interference, since the delay before immediate test is empty and the longer delays are filled with other melodies. Apart from changing two variables in tandem, this design also raises the possibility that subjects will use different strategies than they would when confronted with only one kind of delay, filled or empty. Therefore, Experiments 2 and 3 replicated Experiment 1, but with brief empty and longer filled delays presented to different subjects. Experiment 4 used only empty delays but of a variety of lengths up to 33 sec. Comparison of Experiments 3 and 4 allowed us to assess the degree to which the shift in T/SC versus T/DC performance is due to interference or to the passage of time.

(3) Experiments 1–4 and Dowling (1991b) obtained their results with artificial melodies designed to simplify rhythmic constrasts and vary contour systematically. The question remains, would similar results arise with "real" melodies? Experiments 5 and 6 paralleled Experiments 2 and 3, but with actual melodies familiar or unfamiliar to the subjects.

(4) Following the filled delays of Experiments 1, 3, and 6, the test melody was always transposed in pitch with respect to the initial presentation of the target. This leads to the question, is performance after a filled delay affected by transposition? That is, does the pitch register in which a melody is originally presented serve as an important cue to recognition? Alternatively, if there are confusions on immediate test due to key similarity (as Dowling, 1991b, suggested), are those confusions still present after a brief filled delay, and would they disappear if test stimuli were in the same key as standard stimuli? Experiment 7 replicated Experiment 6, only with T and SC test stimuli presented at the same pitch level as at original presentation.

GENERAL METHOD

The present experiments used the following general design, departures from which are described in the particular Method sections. Most experiments consisted of two sessions: one with tonal stimuli and one with atonal. Each session in the first three experiments consisted of 122 trials of a continuous-running-memory task (Shepard & Teghtsoonian, 1961) made up of 60 intermingled pairs of items and introduced by two buffer items. (Experiment 4 had 90 trials per session.) As shown in Figure 3, each new item could be tested after various delays filled (or not) with intervening items. Each bracket denotes a pair of items. Thus, the melody introduced on Trial 1 was tested on Trial 6, the melody introduced on Trial 2 was tested immediately on Trial 3, and so on. Note that in such a task, subjects respond to every item in the list, including items filling the delay between the introduction of an item and its test. This means that subjects' set in perceiving each item, whether a newly introduced or a test item, is the same. The task also has the motivational advantage that subjects find it more congenial than tasks in which they continually have to "shift gears" in altering their set for a more stressful test sequence.

Each of the 60 pairs consisted of an initial item having a novel contour (different from every other pair in the session) and a test item. In 30 of the pairs, the test item was an exact transposition (T) of the initial item. In the other 30 pairs, the test item was a contour-preserving lure (SC) having the same contour as the initial item but the fourth and fifth notes were altered in pitch from those of an exact transposition by 1.0 to 4.0 semitones. For tonal stimuli, altered pitches remained within the key of the melody. Different-contour (DC) lures were new items that occurred in a temporal position appropriate for a test and that were in the same tonality condition as the initial item being tested. False-alarm rates to DC lures could be collected this way, since subjects responded to each trial as though it were a test trial. A given DC item was assigned as a test of only one initial item. Not all new items functioned as DC items for purposes of data analysis—only those falling in appropriate test positions.

Each of the first cluster of experiments consisted of two sessions of 122 trials each. Tonality condition varied between sessions, so that each subject received a session with tonal melodies and a session with atonal melodies. Approximately half the subjects at each experience level received each order of conditions. The 120 valid trials in a session were introduced by five or six sample stimuli illustrating the contrasts shown in Figure 1, plus two buffer trials containing novel melodies to which subjects responded but that were not scored.

Experiments 5, 6, and 7 used actual melodies in place of artificial ones. The melodies were either familiar to the subjects or not, and the variable familiar versus unfamiliar replaced the variable tonal versus atonal in the design. Twenty-four familiar melodies were chosen as well known to the subjects on the basis of previous

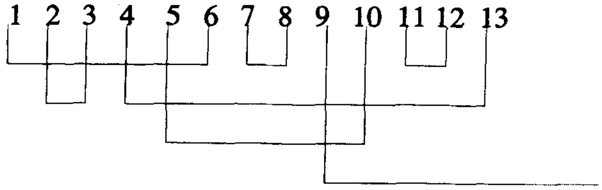

Figure 3. Organization of part of the list in the continuous-running-memory task. Trials 1, 2, 4, 5, 7, 9, and 11 introduce new melodies. Trials 3, 8, and 12 are tests after brief, empty delays of melodies introduced on Trials 1, 7, and 11. Trials 6, 10, and 13 are tests after filled delays of melodies introduced on Trials 1, 5, and 4.

experiments, and 24 stylistically comparable melodies were chosen from McColl and Seeger's (1977) *Travellers' Songs From England and Scotland.* The familiar melodies included such tunes as "Over the Rainbow," "Frosty the Snowman," "Take Me Out to the Ball Game," and "When the Saints Go Marching In," as well as highly familiar folk, nursery, and patriotic tunes. Each of these experiments consisted of a single session of 96 trials in which familiar and unfamiliar melodies were mixed.

The experiments were generally arranged in a 2 experience levels × 3–5 test delays × 2 tonality conditions (tonal, atonal) × 2 test comparisons (T/SC, T/DC) design. All but the first of those variables involved within-groups comparisons. Area under the memory operating characteristic (MOC) was calculated for the T/SC comparison using responses at each confidence level to T comparisons as hit rates and responses at each confidence level to SC lures as false-alarm rates (Banks, 1970). T/DC areas used hit rates to T comparisons and false-alarm rates to DC lures.

Subjects

Undergraduates at the University of Texas at Dallas (mean age 29.5 years) served as part of their course requirements in psychology. Those categorized as musically experienced had more than 2.0 years of explicit musical training ($M = 7.0$ years, $SD = 4.3$ years). Those with less training were categorized as musically inexperienced.

Stimuli

Each stimulus consisted of a seven-note melody presented at a rate of 3.0 notes/second (see Figure 1). The final note of each melody was twice the duration of each of the first six notes. Such melodies group themselves rhythmically as two triplets followed by a longer note. A silent response period of 7.0 sec followed each stimulus in the list. Stimuli were produced on a Roland U-220 synthesizer using its "electric piano" voice, controlled via a MIDI interface by a PC-type computer running Cakewalk software. The stimuli were recorded on tape and presented to subjects via loudspeakers at comfortable levels.

The 120 valid trials in a session consisted of 60 pairs of melodies, such that the first member of each pair introduced a novel melody that differed in contour from all other melody pairs in the list. The 60 contours were chosen from the 64 possible contours of seven-note melodies, omitting the two uniformly rising and falling contours plus two arbitrarily selected contours that had just one contour inflection (reversal of direction). The second member of the pair was a T or SC item and shared the contour of the initial member.

In generating the 60 pairs of melodies, the experimenter first constructed a tonal melody pair for each contour. Tonal melodies conformed to the diatonic major scale and began and ended on the first scale degree (*do*, the tonic). We attempted to construct melodies that would be tonally strong in the sense of Cuddy, Cohen, and Mewhort (1981) by using pitches from the principal triads in the key, with successive notes tending to come from the same triad so that there were fewer underlying triads than there were notes. The triads usually appeared in the order I–IV–V–I. Melodies started and ended on the tonic. Wherever the contour permitted, the melody ended on the same pitch with which it began. Otherwise, it ended with the same pitch class an octave higher or lower. The melodies were also designed to be "melodious" (Pechstedt, Kershner, & Kinsbourne, 1989) in the sense of having relatively narrow pitch intervals and forming relatively coherent gestalts. The second member of SC pairs was formed by altering the pitch of the fourth and fifth notes of the initial melody. This was done in a way that preserved the contour, and usually the harmonic structure, of the original. Thirty melody pairs were randomly assigned to each tonality condition. Atonal melodies were formed by altering their pitch patterns to conform to the atonal "scale" pattern described below.

Atonal melodies began and ended with the same pitch class, but were based on a set of pitch intervals that guaranteed that none conformed to any single tonality. Each atonal melody pair was derived from its tonal counterpart by preserving the "diatonic" intervals of the tonal melody but applying them to a nontonal "scale" pattern. In all but Experiment 2, the nontonal scale pattern changed from pair to pair, being sampled from a set of such scales designed to be as different as possible from any major or minor scale pattern in Western music. To generate this set of nontonal scale patterns, we first generated 35 different patterns of intervals produced by designating seven scale pitches out of the 12 equal-tempered semitones in the octave. Out of those 35 patterns, we found 4 patterns that diverged as much as possible from familiar major and minor scales. Those 4 patterns had the following interval patterns (in semitones) between successive pitches: (3, 3, 1, 1, 1, 2, 1), (3, 1, 1, 2, 2, 2, 1), (3, 1, 2, 1, 2, 1, 2), and (3, 2, 1, 2, 1, 2, 1). We generated 28 possible atonal patterns by taking each of the seven pitches in each of these 4 patterns as the point of origin. Each set of 28 successive pairs in the atonal condition of Experiment 1 was based on a random permutation of those 28 patterns, with each pair based on a different pattern. Thus, each pattern was used two or three times in the session.

The 60 melody pairs of each tonality condition were randomly divided into two groups of 30 pairs, each to be tested with T and SC comparisons. (All melodies were tested with DC comparisons.) Each of those groups of 15 pairs was again divided randomly and assigned to the various delay conditions. The list of 120 trials consisted of five blocks of 24 trials each. The 24 trials in each block consisted of a random order of each of 12 melody pairs assigned to the set of trial types defined by the dimensions of tonality, comparison, and delay.

All comparison members of tonal pairs were in keys moderately distant from those of initial members, having a distance of two, three, or four steps around the circle of fifths in either direction from the origin. This meant that comparison melodies always started and ended on a pitch 2, 3, or 4 semitones above or below that of the initial melody of the pair (3 or 4 in Experiment 2; 2, 3, or 4 in Experiments 1 and 3; and 2 or 3 in Experiment 4). For atonal pairs, this relationship held for the starting and ending notes of the melody. Furthermore, successive trials were in keys (or based on starting notes of atonal melodies) that were 1–4 semitones apart (with keys two to five steps apart around the circle of fifths). Each of the 12 possible keys appeared approximately equally often in the series of 122 trials.

Experiments 5, 6, and 7 used familiar and unfamiliar folk tunes. The first two phrases or so of each melody were used with their natural rhythms for a length of 11–21 notes and were presented at an average tempo of 1.6 notes/second. Melodies averaged about 8.5 sec in length, followed by a response interval of 7.0 sec. Familiar versus unfamiliar replaced tonal versus atonal as a variable, and the two melody types were intermingled in the same session of 96 trials. In other respects, the stimuli and experimental design were the same as in the previous experiments. The unfamiliar songs were chosen from McColl and Seeger (1977) so as to be stylistically similar to the familiar songs. That is, they were unambiguously in a major or minor mode and of fairly regular meter. SC items were produced by changing two pitches in the middle of the song to other diatonic pitches in a way that preserved the contour. In general, we tried to avoid changing strongly accented notes.

For purposes of counterbalancing, different groups received different versions of the list. The main goal of counterbalancing was to ensure that equal numbers of subjects received lists in which a given melody was tested at different delays, so that effects attributed to test delay could not be due to the memorability of particular items. That is, to generate each additional counterbalancing list, the original 60 (or 48) melody pairs were reassigned to delay conditions with the constraint that no pair could be assigned again to a delay condition to which it had already been assigned on previous lists.

Procedure

The subjects were introduced to the experiment by brief explanations of the continuous-running-memory task, the differences among T, SC, and DC test items, and the confidence-level response scale. The experimenter explained the intermingling of new items and test items using a diagram similar to Figure 3. The experimenter informed the subjects that several items might intervene before an item was tested, but that the delay would never be longer than a certain amount (two more than the longest delay being tested in the experiment in question). By means of the familiar tune "Twinkle, Twinkle, Little Star," the experimenter demonstrated the difference between the exact transposition of a melody and an SC imitation. This concept was easy for the subjects to grasp, since exact interval sizes are well remembered for familiar tunes even by inexperienced subjects (Bartlett & Dowling, 1980). The subjects were told to respond positively only to exact transpositions. The experimenter explained the use of the six-category confidence-level scale, with numerical responses from 1 to 6 indicating *very sure new*, *sure new*, *new*, *old*, *sure old*, and *very sure old*. Following those instructions, the experimenter introduced three or four sample trials illustrating the continuous-running-memory task (where appropriate), including the various comparison types (T, SC, and DC). The samples were strongly tonal (or familiar) items with contours not used in the rest of the experiment. The subjects responded during the experiment with response category numbers on a numbered answer sheet.

The subjects performed the 122 (or 96) trials of the session in about half an hour and also completed a brief questionnaire concerning musical experience. On the questionnaire, the subjects were asked to specify ages at which they had had particular musical training and performance experience.

Data Analysis

We report two kinds of data analysis. First, to assess accuracy of memory performance we used area under the MOC as an estimate of unbiased proportion correct where chance is 0.50 (Swets, 1973). Two areas under the MOC were computed for each subject for each tonality × delay combination of conditions. One area compared hit rates to Ts and false-alarm rates to SC lures; the other compared hit rates to Ts and false-alarm rates to DC lures in the same tonality condition. Banks (1970) describes the calculation of area from confidence-level judgments. Areas under the MOC were evaluated by means of analysis of variance (ANOVA).

Second, we examined proportions of hits and false alarms defined by responses of *old*, *sure old*, and *very sure old* (Response Categories 4, 5, and 6) to Ts (hits) and to SCs and DCs (false alarms). Hits and false alarms were evaluated by means of ANOVA.

We also used the criterion split between Response Categories 3 (*new*) and 4 (*old*) to calculate d' and c as measures of performance and response bias, respectively (Macmillan & Creelman, 1991). ANOVAs of d's agreed qualitatively with ANOVAs of area scores. We prefer area under the MOC to d' as a performance measure because (at least for this data) it was more nearly independent of the bias measure c. For example, when d' was correlated with c across subjects within each of the 20 conditions of Experiment 1, the average value of r was .41 (interquartile range .15–.65), while the same correlations between area and c averaged .24 (interquartile range .08–.38). Furthermore, because it makes use of all the response category distinctions used by the subject, area under the MOC also makes more use of the information available in the data and (with "grainy" data such as those of the present experiments) is less sensitive to kinks in the shape of the MOC than is d' (which relies on only one criterion placement).

EXPERIMENT 1

Method

The stimuli and procedure of Experiment 1 were as described above.

Design. Experiment 1 included five delays filled with 0, 1, 2, 4, or 8 intervening items, resulting in an unfilled delay time of 7 sec and filled delay times of 17, 27, 47, or 87 sec. The design involved 2 experience levels × 2 tonality conditions × 5 delays × 2 comparison types, in which only experience was a between-groups variable. Tonality varied between sessions, so that there were 6 trials contributing to each data point for T and SC trials, and 10–12 trials per data point for DC trials.

Subjects. Sixty-five subjects served in Experiment 1; however, 8 had to be dropped from the analysis because of failure to follow directions.

Results

Area scores for Experiment 1 are shown in Figure 4. These data were subjected to a four-way ANOVA: 2 experience levels × 2 tonality conditions × 5 delays × 2 comparison types. The effect of tonality was significant [$F(1,55) = 15.03, p < .001$], with tonal items easier than atonal items (.64 vs. .59). The effect of delay was significant [$F(4,220) = 3.20, p < .02$], with best overall performance at delays of 7 and 47 sec (.64 and .63, respectively) and poorest performance at delays of 17, 27, and 87 sec (.60, .61, and .59, respectively). The effect of comparison type was significant [$F(1,55) = 23.27, p < .001$], with T/SC comparisons easier (.63) than T/DC comparisons (.60). The interaction of tonality × comparison type was significant [$F(1,55) = 10.61, p < .01$], with the effects of tonality more pronounced for T/SC

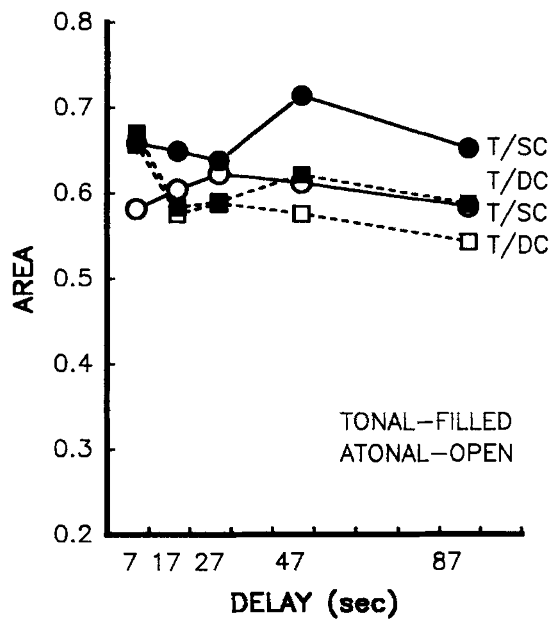

Figure 4. Areas under the MOC for the tonal (filled symbols) and atonal (open symbols) conditions of Experiment 1. Circles denote T/SC discrimination, and squares denote T/DC discrimination.

comparisons than for T/DC comparisons (see Table 1). The interaction of delay × comparison type was significant [$F(4,220) = 10.12$, $p < .001$], with performance starting out strong for T/DC comparisons but falling below that of T/SC comparisons over the longer filled delays. T/SC discrimination remained strong over the longest delay of 87 sec. That interaction is shown in Figure 5. The interaction of delay × comparison type × experience was significant [$F(4,220) = 2.85$, $p < .05$], in which performance for the inexperienced subjects on T/SC comparisons peaked more sharply and on T/DC comparisons fell off more rapidly than for the experienced subjects. No other main effects or interactions were significant.

In the ANOVA on proportions of hits and false alarms, the results (shown in Figure 6) were qualitatively similar to those of the analysis of area scores. The main effects of delay [$F(4,220) = 8.33$, $p < .001$] and comparison type [$F(2,110) = 146.04$, $p < .001$] were significant. The interactions of tonality × comparison type [$F(2,110) = 11.27$, $p < .001$], delay × comparison type [$F(8,440) = 2.08$, $p < .05$], and tonality × delay × comparison type [$F(8,440) = 2.40$, $p < .02$] were also significant. No other main effects or interactions were significant.

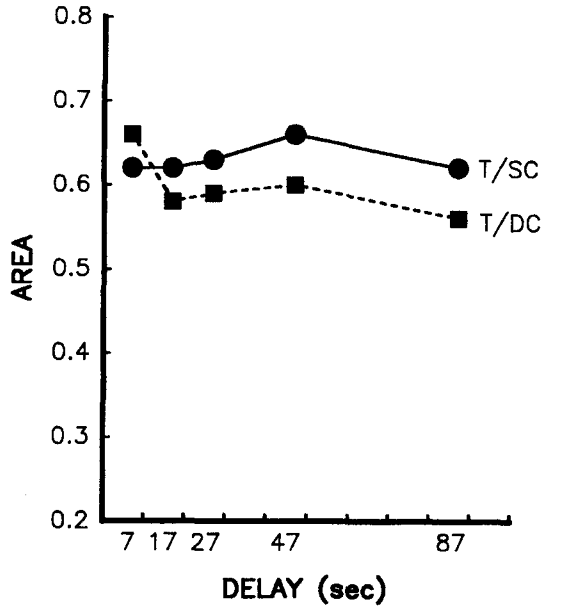

Figure 5. Areas under the MOC for Experiment 1 collapsed across tonality conditions. Circles denote T/SC discrimination, and squares denote T/DC discrimination.

Discussion

It is clear from Figure 5 that only after the brief empty delay was T/DC superior to T/SC performance. The T/DC–T/SC shift was even stronger here than in Dowling's (1991b) study (Figure 2). For all the filled delays from 17 to 87 sec, T/SC performance remained superior to T/DC performance and, in fact, peaked at 47 sec. Given the large physical differences between T and DC patterns, and the subtle differences between T and SC patterns, this shift over time is surprising and in need of explanation.

The superior performance of T/DC discrimination over T/SC discrimination on immediate test has been obtained before and can be attributed at least in part to the salience of contour and confusions due to contour similarity of Ts and SCs, coupled with contour differences of DCs. We can, however, reject Dowling's (1991b) hypothesis tying those confusions to key similarity, since false alarms to SCs after the empty delay declined at least as steeply for atonal stimuli as for tonal stimuli (Figure 6). We can also reject explanations that attribute the decline in T/DC performance to a rapid decline in memory for contour, whether through forgetting or through declining usefulness (see Dowling & Bartlett, 1981), since false alarms to DCs remained at about the same level throughout the delays tested (Figure 6).

The data in Table 1 show that the pitch intervals of tonal melodies (that conform to the invariants of the tonal scale pattern) are more accurately encoded than those of atonal melodies (that do not). Those data also show that contours (which are independent of tonality) are encoded equally well for tonal and atonal melodies.

Both the shift over time in T/DC and T/SC performance and the interaction of those measures with tonality support the proposition that contour information and pitch-interval-pattern information function differently in recognition. Because of that difference, the subjects might have been drawn by the mixing of filled and unfilled delays into using mixed strategies, attempting to boost T/DC performance for immediate comparisons and T/SC performance for delayed comparisons. The subjects' performance might have been different if the task included only empty or filled delays. Therefore, we designed Experiments 2, 3, and 4 to test recognition with only empty or only filled delays. Experiments 2 and 3 formed a complementary pair, essentially replicating the brief empty and longer filled delays of Experiment 1, respectively, but with different subjects performing the two types of task. Experiment 4 tested a set of longer empty delays than those in Experiments 1 and 2.

Table 1
Areas Under the MOC for Transpositions (T) Versus Same-Contour (SC) and Different-Contour (DC) Lures in Tonal and Atonal Sessions in Experiments 1 and 3

Comparison	Tonal	Atonal	*M*
	Experiment 1		
T/SC	.66	.60	.63
T/DC	.61	.59	.60
	Experiment 3		
T/SC	.63	.59	.61
T/DC	.57	.56	.57

EXPERIMENT 2

Experiment 2 contained only immediate tests after empty delays of 7 sec. This was intended to induce sub-

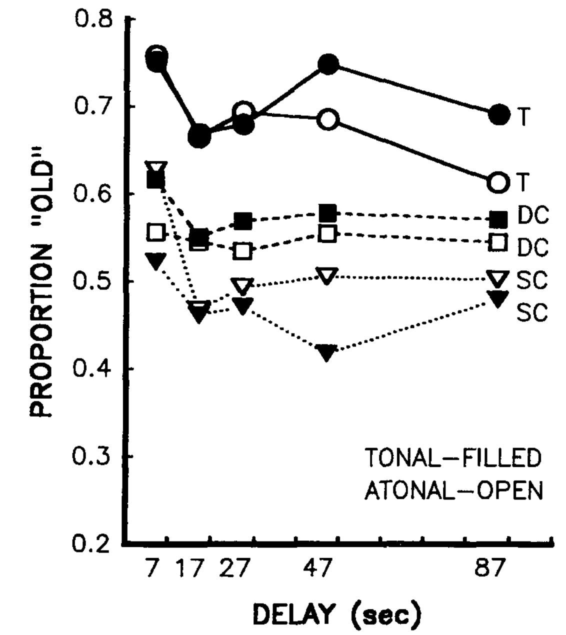

Figure 6. Proportion "old" judgments for the tonal (filled symbols) and atonal (open symbols) conditions of Experiment 1. Circles denote hit rates to Ts; triangles, false-alarm rates to SC lures; and squares, false-alarm rates to DC lures.

jects to adopt a strategy appropriate to unfilled brief delays, without simultaneously coping with the burden on memory imposed by longer filled delays. We wished to see whether, with only immediate tests to plan for, subjects would behave in the same way as when immediate tests were mixed with delayed tests. In particular, we wanted to see if subjects would adopt a strategy emphasizing contour recognition and minimizing the effects of tonality.

Method

Design. Experiment 2 was exactly like Experiment 1, except that the subjects were instructed to respond to each stimulus only with regard to the stimulus immediately preceding it, and the results were scored accordingly. This resulted in a 2 experience levels × 2 tonality conditions × 2 comparison types design, since there was only one delay of 7 sec containing no intervening items.

Subjects. Twenty-eight subjects served in Experiment 2; however, 1 had to be dropped from the analysis because of failure to follow directions.

Procedure. The procedure was the same as that for Experiment 1, except for the differences in instructions and scoring described above.

Results

The ANOVA on area scores from Experiment 2 showed only a main effect of comparison type [$F(1,25) = 64.85$, $p < .001$], with T/DC comparisons easier (.78) than T/SC comparisons (.58). Those results can be seen in Figure 7. The results of the ANOVA on proportions of hits and false alarms were similar, showing only a main effect of comparison type [$F(2,50) = 59.95$, $p < .001$]. Proportions of *old* responses to Ts, SCs, and DCs were .70, .58, and .32, respectively.

Discussion

A comparison of the results of Experiments 1 and 2 shows that when the burden of planning for delayed tests is removed, subjects' T/DC performance improves by about 10 percentage points. Subjects' tendency to concentrate on contour discrimination on immediate tests, evident in Experiment 1, becomes even more pronounced when only immediate tests are required. This improvement in T/DC discrimination was accomplished with very little cost to T/SC discrimination.

A consequence of the strategy shift between Experiments 1 and 2 is the disappearance of tonality effects. Apparently, tonality is important only when subjects perceive the encoding of intervallic detail as important. When subjects concentrate on contour discrimination (as in Experiment 2), tonality becomes less relevant to performance. This agrees with the results of Experiment 1 concerning the lack of effect of tonality on T/DC performance (see Table 1).

EXPERIMENT 3

Experiment 3 served as the complement of Experiment 2 in utilizing only filled intervals. Here, we wanted to see whether we would find the same pattern of performance with filled intervals as in Experiment 1, even when subjects no longer had to perform the immediate recognition task in the same experiment.

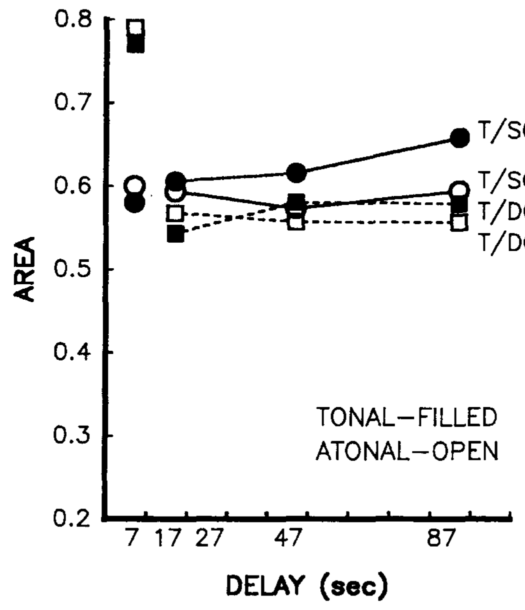

Figure 7. Areas under the MOC for the tonal (filled symbols) and atonal (open symbols) conditions of Experiments 2 and 3. Circles denote T/SC discrimination, and squares denote T/DC discrimination. The results of Experiment 3 are connected with lines, whereas those of Experiment 2 are isolated at the left of the figure.

Method

Design. There were three filled delays in Experiment 3, with one, four, or eight intervening items, lasting 17, 47, or 87 sec, respectively. The design comprised 2 experience levels × 2 tonality conditions × 3 delays × 2 comparison types, in which only experience was a between-groups variable. There were 10 trials per data point for T and SC stimuli and 18–22 trials per data point for DC stimuli.

Stimuli. The stimuli were as described in the General Method section. Three counterbalanced lists were used with approximately equal numbers of experienced and inexperienced subjects. Particular melodies were included in different delay conditions in the different lists.

In Experiment 3, we returned to a simpler method for generating atonal melodies than that used in Experiment 1. Here, one invariant set of intervals underlay all the atonal patterns. The pitches of that set were arranged with the following semitone intervals between them, starting with the beginning and ending note (1, 3, 1, 1, 3, 2, 1), or, in pitch classes beginning on C (C, D♭, E, F, G♭, A, B). The two clusters with one-semitone spacing (B-C-D♭ and E-F-G♭) ensured that the resultant patterns could not conform to a tonal scale.

Subjects. Thirty-eight subjects served in Experiment 3; however, 2 had to be dropped from the analysis because of failure to follow directions.

Procedure. The procedure for Experiment 3 was the same as that of Experiment 1.

Results

Area scores from Experiment 3 are shown in Figure 7. These data were subjected to a four-way ANOVA: 2 experience levels × 2 tonality conditions × 3 delays × 2 comparison types. The effect of comparison type was significant [$F(1,34) = 21.12$, $p < .001$], with T/SC comparisons easier (.61) than T/DC comparisons (.57). The only other significant effect was that of the tonality × comparison type interaction [$F(1,34) = 6.23$, $p < .02$] shown in Table 1, in which tonality conferred an advantage on T/SC comparisons but not on T/DC comparisons.

The ANOVA on hits and false alarms showed comparable results, with a significant effect of comparison type [$F(2,68) = 67.30$, $p < .001$] and a significant interaction of tonality × comparison type [$F(2,68) = 5.04$, $p < .01$]. No other effects were significant. As might be expected given the area scores in Figure 7 and the lack of effects of delay in the ANOVA, hits and false alarms remained relatively constant across the delays tested: hits were in the range .64–.67, false alarms to SC lures .47–.50, and false alarms to DCs .53–.56.

Discussion

Comparison of the combined results of Experiments 2 and 3 (Figure 7) with the results of Experiment 1 (Figure 4) discloses much the same pattern: T/DC performance was better than T/SC performance after a brief empty delay, but the reverse was true after longer filled delays. The largest effect of separating empty and filled delays appears to be the improvement in T/DC discrimination after the empty delay in Experiment 2. We attribute this improvement to a shift in strategy to concentrate on contour recognition. It is interesting that this shift did not incur much cost in terms of a decline in T/SC discrimination, which, in Experiments 1 and 2, remained close to .60. Furthermore, when only filled delays were tested in Experiment 3, performance was about the same as, or perhaps a little lower than, in Experiment 1. (This is most clearly seen in Table 1.)

The results of Experiments 2 and 3 clarify the effects of tonality that were present in Experiment 1. When the subjects concentrated on T/DC discrimination in the immediate recognition task of Experiment 2, the effects of tonality disappeared. When the subjects concentrated on T/SC discrimination with the filled delays of Experiment 3, the same tonality × comparison interaction appeared as that in Experiment 1, with tonality affecting T/SC discrimination but not T/DC discrimination (Table 1). Again, this suggests the importance of tonality to the encoding of fine intervallic detail in melodies, so that that information can be used when tests occur after filled delays. (There was no main effect of tonality in Experiment 3, but that was probably due to the weaker manipulation of tonality than in Experiment 1.[2])

It is also noteworthy that in Experiment 3 there were no effects of delay. Performance, while not far above chance, remains about constant over the 87-sec delay, with even a suggestion that tonal T/SC discrimination may still be rising at the end of that period. There is surprisingly little forgetting, in spite of the interference in the filled delays.

EXPERIMENT 4

Given the results so far, the question remains whether the pattern of results for filled delays seen in Experiments 1 and 3 would hold for comparable delays in which there was no distracting material between the introduction of a melody and its test. Therefore, in Experiment 4, we tested for recognition after empty delays of 7, 12, and 33 sec. In particular, we wanted to find out whether T/DC discrimination would remain superior to T/SC discrimination as long as there were no intervening melodies or whether the shift seen in Figure 5 was due to the mere passage of time. We were also curious whether tonal T/SC discrimination would improve during the first 30 sec (as it did in Experiment 1) in the absence of intervening distractors.

Method

Design. Experiment 4 was similar to Experiment 1, except that there were three empty delays of 7, 12, and 33 sec. This resulted in a 2 experience levels × 2 tonality conditions × 3 delays × 2 comparison types design, in which only experience was a between-groups variable. Each session consisted of 90 valid trials preceded by three buffers. Reducing the number of trials from 120 to 90 was necessitated by the additional time added to the session by the longer empty delays. We did not include even longer delays because pilot work convinced us they would be so boring for the subjects that such a study would not be comparable in motivational aspects to the preceding three experiments. For example, adding 18 trials (6 each of T, SC, and DC tests) at 60 sec each would add 18 min to the session, all in empty time. DC trials consisted of first members of pairs used for T and SC trials. There were 6 trials contributing to each data point for each type of trial.

Subjects. Fifty-six subjects served in Experiment 4; none had to be dropped from the analysis.

Procedure. The procedure was the same as for Experiment 3, and the subjects were instructed to respond to each stimulus only in comparison to the immediately preceding stimulus.

Results

The ANOVA on area scores from Experiment 4 showed main effects of delay [$F(2,108) = 10.58$, $p < .001$] and comparison type [$F(1,54) = 32.00$, $p < .001$]. Long delays were more difficult than short delays, with performance going from .67 to .60 between 7 and 33 sec. T/DC comparisons were easier (.66) than T/SC comparisons (.59). The interaction of delay × comparison type was significant [$F(2,108) = 7.79$, $p < .001$] and is shown in Figure 8. Performance on T/DC comparisons declined more sharply with time than did performance on T/SC comparisons. There were no other main effects or interactions.

The results of the ANOVA on proportions of hits and false alarms were similar, showing only main effects of delay [$F(2,108) = 18.50$, $p < .001$] and comparison type [$F(2,108) = 61.44$, $p < .001$], and a delay × comparison type interaction [$F(4,216) = 10.93$, $p < .001$]. Those data are shown in Figure 9.

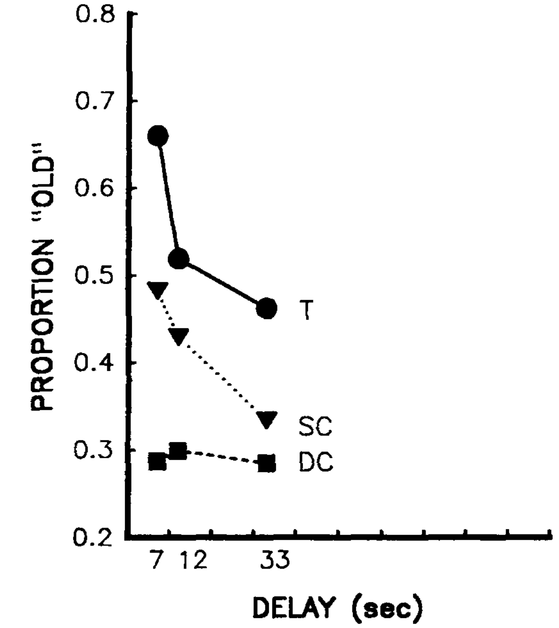

Figure 9. Proportion "old" judgments collapsed across tonality conditions for Experiment 4. Circles denote hit rates to Ts; triangles, false-alarm rates to SC lures; and squares, false-alarm rates to DC lures.

Discussion

In Experiment 4, as in Experiment 2, when subjects did not confront tests after filled delays, they focused their strategies on discriminating contour differences. Also, as in Experiment 2, this strategy was associated with a lack of effect of tonality. However, it is clear from Figure 9 that the subjects were not simply doing contour discrimination. The subjects performed better than chance on T/SC discrimination and, at the shortest delay, performed just as well as in Experiment 1. And as in Experiment 1, although false alarms to DC lures remained constant over the 33-sec empty delay, hits declined severely. Furthermore, between 12 and 33 sec, false alarms to SC lures declined more sharply than did hits, with the result that T/SC performance was closer to T/DC performance at 33 sec than at 7 or 12 sec (Figure 8).

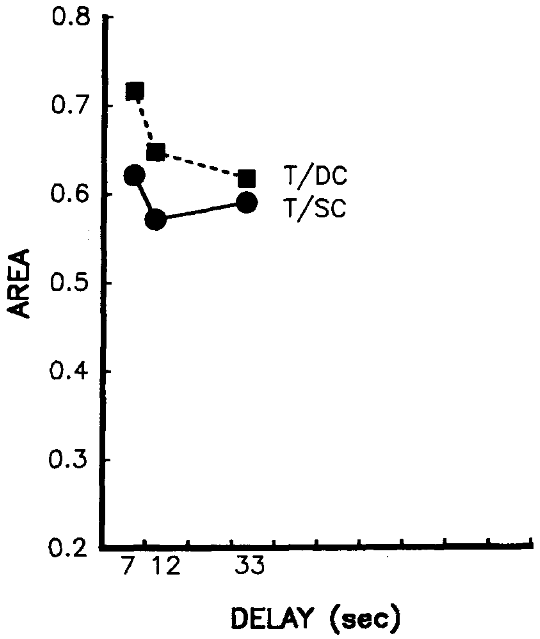

Figure 8. Areas under the MOC for Experiment 4 collapsed across tonality conditions. Circles denote T/SC discrimination, and squares denote T/DC discrimination.

In Experiment 4, we again see changes over time in T/DC performance relative to T/SC performance, with the former declining much more rapidly than the latter. It seems unlikely to us, however, that, in the absence of intervening stimuli, T/SC would surpass T/DC performance. But here, as in Experiments 1–3, contour recognition and pitch-interval pattern recognition followed different patterns of performance over time, and, in all the experiments, we see the same shift of relative strength in favor of T/SC performance.

In Experiments 1–4, we used isochronous seven-note melodies. In Experiments 5–7, we used actual folk melodies. There are good reasons for using actual melodies. First, we need them if we wish to generalize to people's memory for melodies in everyday musical experience. Second, though the contours of the artificial melodies of Experiments 1–4 are carefully controlled so as to be different from one another, they lack the rhythmic differentiation of the melodic–rhythmic contours (Monahan & Carterette, 1985) of actual melodies. The informational impoverishment of the contours of artificial melodies might conceivably be responsible for the relatively poor T/DC performance with delayed tests in Experiments 1–4.

There are at least two ways in which the simplification of contour could affect the relationship between T/SC and T/DC performance. First, while T and SC stimuli have an obvious similarity to the contour just presented on immediate test, with increasing delay, contours should be less effective in reminding the listener of the relevant standard stimulus. This decline in effectiveness should be steeper with simplified contours than with richer natural contours. The simplified contours of isochronous melodies will be less effective cues in reminding listeners of relevant potential matches than will the informationally richer melodic–rhythmic contours of actual melodies. Thus, positive responses to Ts and SCs should decline steeply, whereas DC false alarms will remain roughly constant.

Second, though informationally impoverished contours may sometimes serve as effective retrieval cues for potentially relevant standard stimuli, they should be less effective in leading to the rejection of DC lures, since there are fewer features in which they could differ from a new contour when a match is attempted. (In this case, we assume that retrieval of potential matches is made on the basis of some global sense of familiarity, whereas acceptance or rejection of the candidates retrieved is made on the basis of feature matches and mismatches, especially the latter.) Here, we must complicate our theorizing by noting that simply adding contour information does not lead to improved T/DC discrimination. Edworthy (1985) found that lengthening isochronous melodies (thus adding contour information) led to a decrement in performance. What is important about actual melodies is that the information is added along a new dimension (i.e., rhythm), making it easier to use in working memory (as Miller, 1956, suggested). Therefore, with these natural melodies with their melodic–rhythmic contours, we expect T/DC discrimination to be better than with isochronous melodies.

For these reasons, we expected an improvement in T/DC discrimination when recognition after filled delays is tested with actual melodies rather than with artificial melodies.

EXPERIMENT 5

Experiments 5 and 6 paralleled Experiments 2 and 3, but instead of artificially constructed melodies used actual melodies, mostly from folksongs, that differed in familiarity.

Method

Design. Experiment 5 was similar to Experiment 4, except that, in Experiment 5, we used actual melodies with only a brief empty delay of 7.0 sec, and familiarity replaced tonal coherence as a variable in the design. The stimuli consisted of 48 pairs of melodies generated from 24 familiar and 24 unfamiliar folk melodies. There were 48 trials in which each melody was tested immediately with one of the three comparison types: T, SC, or DC.

Subjects. Eighteen subjects served in Experiment 5; none were excluded from the analysis.

Procedure. The procedure of Experiment 5 was the same as that for Experiment 4, except that there was only one empty delay of 6 sec.

Stimuli. The stimuli consisted of the first phrase or two of 24 familiar tunes and 24 unfamiliar folksongs, as described in the General Method section.

Results

Area scores for Experiment 5 are shown in Figure 10. These data were subjected to a three-way ANOVA: 2 experience levels × 2 familiarity conditions × 2 comparison types. The effect of familiarity was significant [$F(1,16) = 83.75, p < .001$], with familiar items easier than unfamiliar items (.94 vs. .67). The effect of comparison type was significant [$F(1,16) = 86.15, p < .001$], with T/DC comparisons easier (.91) than T/SC comparisons (.70). The interaction of familiarity × comparison type was significant [$F(1,16) = 38.26, p < .001$], with the difference between T/SC and T/DC performance greatest for unfamiliar items, as shown in Figure 10. There were no other main effects or interactions.

The ANOVA on proportions of hits and false alarms showed qualitatively similar results, including a main effect of familiarity [$F(1,16) = 6.16, p < .05$], a main effect of comparison type [$F(2,32) = 103.04, p < .001$], and an interaction between them [$F(2,32) = 51.50, p < .001$].

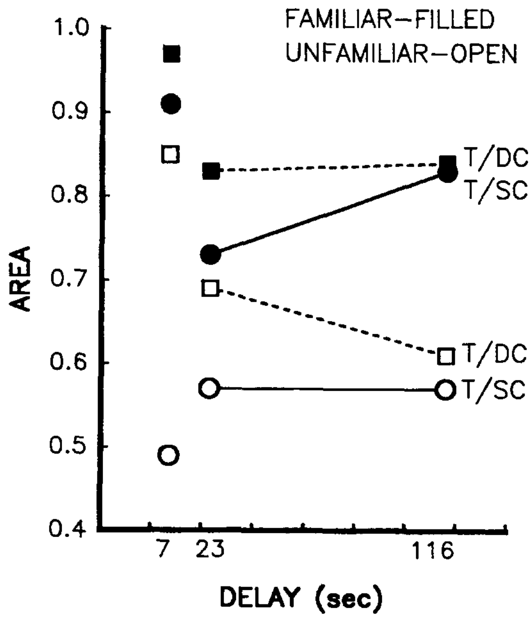

Figure 10. Areas under the MOC for the familiar (filled symbols) and unfamiliar (open symbols) conditions of Experiments 5 and 6. Circles denote T/SC discrimination, and squares denote T/DC discrimination. The results of Experiment 6 are connected with lines, whereas those of Experiment 5 are isolated at the left of the figure.

EXPERIMENT 6

Method

Design. Experiment 6 served as the complement to Experiment 5, providing tests after filled delays using the same 48 actual melodies. Like Experiment 3, Experiment 6 involved a continuous-running-memory task with two filled delays containing one or seven intervening items. Since the stimuli averaged 8.5 sec in length and the interstimulus response interval was 7.0 sec, the filled delays were 22.5 and 115.5 sec long. Recognition was tested with one of the three comparison types: T, SC, or DC. In the counterbalancing scheme, 12 of each kind of melody were tested at the short delay for one group of subjects and at the long delay for the other group.

Subjects. Thirty-one subjects served in Experiment 6; none were excluded from the analysis.

Procedure. The procedure of Experiment 6 was the same as that of Experiment 3, with 96 trials in one session.

Stimuli. The stimuli were the same as those in Experiment 5.

Results

Area scores for Experiment 6 are shown in Figure 10. These data were subjected to a three-way ANOVA: 2 experience levels × 2 familiarity conditions × 2 delays × 2 comparison types. The effect of experience was significant [$F(1,29) = 9.03$, $p < .01$], with experienced subjects performing better than inexperienced ones (.77 vs. .66). The effect of familiarity was significant [$F(1,29) = 101.80$, $p < .001$], with familiar items easier than unfamiliar items (.80 vs. .61). The effect of comparison type was significant [$F(1,29) = 12.81$, $p < .001$], with T/DC comparisons easier (.75) than T/SC comparisons (.68). The interaction of delay × experience was significant [$F(1,29) = 4.90$, $p < .05$], with inexperienced subjects improving slightly and experienced subjects declining slightly over time. The interaction of familiarity × delay was significant [$F(1,29) = 6.93$, $p < .02$], with performance increasing with delay for familiar items (.78 to .83) and decreasing for unfamiliar items (.63 to .59). The interaction of delay × comparison type was significant [$F(1,29) = 16.19$, $p < .001$], with overall T/SC performance improving with increased delay and T/DC performance declining (see Table 2). There were no other main effects or interactions.

The ANOVA on proportions of hits and false alarms showed qualitatively similar results, including a main effect of familiarity [$F(1,29) = 8.09$, $p < .01$], a main effect of comparison type [$F(2,58) = 80.27$, $p < .001$], and interactions of item type with experience [$F(2,58) = 5.37$, $p < .01$], with familiarity [$F(2,58) = 43.99$, $p < .001$], and with delay [$F(2,58) = 6.94$, $p < .01$].

Discussion

What is most interesting in the results of Experiments 5 and 6 is performance with unfamiliar melodies. Comparison of those results (Figure 10) with those for tonal melodies in Experiments 2 and 3 (Figure 7) provides us with a means of assessing the effects of using real melodies instead of artificial ones. As expected, T/DC performance was better with real melodies, showing improvements of the order of 10 percentage points at delays of 7 and 23 sec. However, that advantage gradually disappeared, so that after 116 sec, T/DC performance for both types of melodies hovered around .60. T/SC performance was somewhat worse with real melodies than with artificial ones, beginning at chance when tested immediately and rising to close to .60 after the filled delays. The pattern of falling T/DC performance and rising T/SC performance held for both sets of experiments, though with real melodies T/DC performance was sufficiently strong so that it was never surpassed by T/SC performance.

Table 2
Areas Under the MOC for Transpositions (T) Versus Same-Contour (SC) and Different-Contour (DC) Lures for Short and Long Delays in Experiments 6 and 7

Comparison	Short	Long	*M*
	Experiment 6		
T/SC	.65	.70	.68
T/DC	.76	.73	.75
	Experiment 7		
T/SC	.67	.71	.69
T/DC	.76	.74	.75

In both Experiment 3 and Experiment 6 (Figures 7 and 10), there was a slight tendency for T/SC performance overall to rise with increasing delay (see Table 2) and definitely not to decline. This is contrary to what one usually finds in memory experiments. This led us to think that some process may have been depressing performance at the earlier filled delays. For example, the confusion that causes T/SC performance to be near chance after the brief empty delay may still operate to some extent after the first filled delay. To test for this possibility, in Experiment 7, we replicated Experiment 6 but with T and SC comparison stimuli in the same key and pitch register as their corresponding standards. If confusion in judging a transposition soon after hearing a standard melody was operating, then T/SC performance for the 23-sec filled delay should improve, but not after the 116-sec delay.

An alternative possibility, compatible with the possibility just described, is that pitch register itself can serve as a retrieval cue. In that case, presenting T and SC comparisons in the same key as their standards, while DC comparisons remain in different keys (and hence at different pitch levels), should enhance T/DC performance by making DCs easier to reject.

EXPERIMENT 7

Method

Design, Stimuli, and Procedure. Experiment 7 was identical to Experiment 6, except that comparison melodies were in the same key and at the same pitch level as corresponding standard melodies.

Subjects. Seventeen subjects served in Experiment 7; none were excluded from the analysis.

Results

The area scores from Experiment 7 closely paralleled those from Experiment 6 and were subjected to a three-way ANOVA, as in Experiment 6. There were fewer significant interactions than in Experiment 6. The effect of experience was significant [$F(1,15) = 16.13, p < .01$], with experienced subjects performing better than inexperienced subjects (.80 vs. .66). The effect of familiarity was significant [$F(1,15) = 28.35, p < .001$], with familiar items easier than unfamiliar items (.80 vs. .64). The interaction of delay × comparison type was significant [$F(1,15) = 4.56, p < .05$], and it corresponded closely to that found in Experiment 6 (see Table 2). There were no other main effects or interactions.

The ANOVA on proportions of hits and false alarms showed results consistent with the above, including a main effect of familiarity [$F(1,15) = 6.47, p < .05$], a main effect of comparison type [$F(2,30) = 36.15, p < .001$], and interactions of item type with experience [$F(2,30) = 4.10, p < .05$], with familiarity [$F(2,30) = 7.56, p < .01$], and with delay and experience [$F(2,30) = 3.34, p < .05$]. There was also an interaction of familiarity × delay [$F(1,15) = 6.99, p < .02$], in which positive responses decreased over time for familiar items and increased for unfamiliar items.

Discussion

There was no tendency for T/SC performance at 23 sec to improve when SC comparisons were presented in the same key as standards. In particular, SC false-alarm rates to unfamiliar melodies in Experiment 7 remained within 3 percentage points of those in Experiment 6. Therefore, the failure to decline of T/SC performance over time does not seem to be due to confusions present at early filled tests that are later dispelled.

On the other hand, it does appear that the subjects were able to use key and pitch level as a cue in rejecting DC lures, especially for inexperienced subjects judging unfamiliar melodies. This is the principal source of the significant item type × delay × experience interaction in the analysis of hits and false alarms. Inexperienced subjects went from a DC false-alarm rate of .34 after 116 sec in Experiment 6 to one of .15 in Experiment 7.

GENERAL DISCUSSION

These experiments on the time course of recognition of novel melodies demonstrate different ways in which contour information and pitch-interval pattern information contribute to recognition. We took T/DC discrimination as an index of the utilization of contour information and T/SC discrimination as an index of the utilization of pitch-interval pattern information. As in previous research, we found that contour dominated recognition following brief empty delays, and pitch-interval pattern dominated recognition after filled delays. This pattern of results essentially replicated the puzzling shift in T/DC and T/SC performance across a filled delay found in earlier studies (DeWitt & Crowder, 1986; Dowling, 1991b; Dowling & Bartlett, 1981). That shift appeared even more dramatically in the present Experiments 1, 2, and 3, in which T/SC performance actually surpassed T/DC performance following filled delays and remained strong even after 87 sec. With real melodies, whose richer melodic–rhythmic contours boosted T/DC performance, T/SC performance did not surpass T/DC (Experiments 5, 6, and 7); the pattern of T/DC decline and T/SC improvement over time remained the same, however. That pattern, but with T/SC performance remaining roughly constant, also held when only empty delays were tested (Experiment 4).

A clear conclusion emerges from the present experiments concerning the role of tonality in memory for melodies. Tonality was not an important factor when the subjects concentrated on T/DC discrimination following the empty delays of Experiments 2 and 4. Tonality was important only in Experiments 1 and 3, where filled delays were involved. There, the principal effect of tonality was to facilitate T/SC discrimination for tonal items, as seen in Table 1. It seems plausible that the pitch patterns of tonal melodies that conform to the invariant pitch patterns of familiar musical keys should be more easily encoded than the pitch patterns of atonal melodies that clearly do not conform to any particular pattern.

We believe that the present results contain useful suggestions concerning the role of automatic and controlled processes in perception and memory for melodies. Automatic processes can be carried out without placing much demand on the information-processing capacity of the system and so can be executed in parallel with other tasks, whereas controlled processes do make demands on the system's capacity (Schneider & Shiffrin, 1977). Automatic processes are carried out without the subject's control and often without the subject's awareness; controlled processes usually require conscious attention. As a result of the relative demands they make on the information-processing capacity of the system, controlled processes tend to interfere with each other when an attempt is made to carry them out simultaneously, whereas automatic processes can be carried out free of interference from other automatic processes or controlled processes (cf. Brooks, 1967, 1968).

We believe the evidence suggests that the encoding of contours is largely a controlled process, and the encoding of pitch-interval pattern information is largely automatic. The critical evidence for this involves interference, though other converging evidence will be cited below. The present experiments show that T/DC discrimination is subject to interference from concurrent tasks, and T/SC discrimination is not. First, in going from Experiment 2 (Figure 7) to Experiment 1 (Figure 4), the main difference is the addition to the immediate comparison task of the task of making comparisons spanning up to 1.5 min of filled delays. The necessity of making these delayed judgments (themselves involving T/DC comparisons) brings about a 10% drop in T/DC performance on the immediate task. Second, in going from Experiment 3 to Experiment 1 (again, Figures 7 and 4), the difference is the addition of an immediate judgment task to the series of delayed comparisons. This addition has virtually no effect on T/SC discrimination on the delayed trials, or if anything leads to an improvement. It seems reasonable to take this asymmetry of effects of interference as an indication that encoding and remembering contour information involves controlled processes, whereas encoding and remembering pitch-interval-pattern information involves automatic processes.

Other observations in the present experiments support the distinction between contour and interval-pattern processing in terms of controlled and automatic processes. First, the proposition that interval-pattern processing is automatic is consistent with the observation that T/SC performance often improved over time across filled delays during which the subject was making other judgments. In fact, T/SC performance did not appear to improve across empty delays (Experiment 4, Figure 8). This converges with the results of Abdi, Piat, and Dowling (1994), who had subjects remember phrases of Schubert songs. They found that certain interfering tasks (e.g., counting backwards) interposed between the introduction of a melody and its test led to dramatically better T/SC performance than did, for example, attempts at active rehearsal by the subjects.

Second, while they are quite aware of doing the contour-discrimination task, subjects are generally unaware of their ability to do T/SC discriminations. At the end of the experimental session, the subjects often expressed the feeling that they were purely guessing on T/SC comparisons of unfamiliar melodies, although, on the average, they were performing well above chance. Again, this is consistent with the notion that T/SC discrimination involves automatic processes.

The present results converge with those of Bartlett, Halpern, and Dowling (1993) comparing young and elderly nonmusicians and musicians on recognition memory tasks requiring T/SC and T/DC discrimination. Those experiments led to the conclusion that T/DC discrimination, as an explicit cognitively controlled task involving working memory (Baddeley, 1990), was more affected by age, whereas T/SC discrimination, as an implicit procedural task not drawing on working-memory capacity, was more affected by expertise.

There is one remaining issue involving an apparent puzzle in subjects' ability to recognize the pitch-interval pattern of melodies in cases where they cannot recognize the contour. If the pitch-interval pattern is represented as something like a sequence of signed intervals in semitones (as described by Dowling, 1978), then why isn't contour information immediately accessible in that representation, as simply the sequence of signs? The answer is that the superiority of T/SC discrimination over T/DC discrimination after filled delays provides one more piece of evidence that pitch-interval information is not represented as a sequence of signed intervals, but rather as a sequence of abstract representations of pitch classes. That is, the start of "Twinkle, Twinkle, Little Star" is represented not as (0, +7, 0, +2, 0, −2), but rather as something like (*do*, *do*, *sol*, *sol*, *la*, *la*, *sol*), where the *do* is moveable—assignable to a new tonic in transposition. A considerable amount of converging evidence for this conclusion is reviewed elsewhere (Dowling, 1991a, pp. 54f). Briefly, that evidence shows that (1) intervals are much easier to retrieve with reference to familiar melodies than vice versa, (2) dynamic tendencies of pitches in a tonal context are properties of pitch classes, not of intervals—that is, inverted intervals having the same pitch classes, such as thirds and sixths, are highly similar (Balzano & Liesch, 1982)—(3) the foregoing leads to the conclusion that the tonal hierarchy of functions of pitches in a tonal context (Krumhansl, 1990) is best defined on pitch classes and not intervals, and (4) the interval pattern of a tonal sequence can be destroyed either by interleaving distractor tones among its pitches or by scrambling its pitches into several octaves, and the sequence remains recognizable in cued recognition (Dowling, 1984; Dowling, Lung, & Herrbold, 1987). For all these reasons, it seems very likely that the information that serves as the basis for T/SC discrimination is represented as sets of abstract pitch classes in something like a moveable-*do* system.

In summary, these experiments on the early time course of melody recognition show that on immediate test after empty delays contour information dominates performance. However, T/DC discrimination fades rapidly, even for real melodies with rich melodic–rhythmic contours. Pitch pattern information, in contrast, dominates performance after filled delays, and T/SC discrimination remains strong for periods as long as 2 min. Tonality appears to be important to the encoding of pitch pattern information. Contour information appears to be under explicit cognitive control, whereas pitch pattern information seems to be represented implicitly and processed automatically.

REFERENCES

Abdi, H., Piat, F., & Dowling, W. J. (1994). Interference in the early time course of memory for melodies. In I. Deliège (Ed.), *Proceedings of the Third International Conference on Music Perception and Cognition* (pp. 249-250). Liege, Belgium: ESCOM.

Attneave, F., & Olson, R. K. (1971). Pitch as medium: A new approach to psychophysical scaling. *American Journal of Psychology*, **84**, 147-166.

Baddeley, A. (1990). *Human memory*. Boston: Allyn & Bacon.

Balzano, G. J., & Liesch, B. W. (1982). The role of chroma and scalestep in the recognition of musical intervals in and out of context. *Psychomusicology*, **2**, 3-31.

Banks, W. P. (1970). Signal detection theory and human memory. *Psychological Bulletin*, **74**, 81-99.

Bartlett, J. C., & Dowling, W. J. (1980). Recognition of transposed melodies: A key-distance effect in developmental perspective. *Journal of Experimental Psychology: Human Perception & Performance*, **6**, 501-515.

Bartlett, J. C., Halpern, A., & Dowling, W. J. (1993). *Recognition memory and source monitoring for music in normal aging and Alzheimer's disease*. Manuscript submitted for publication.

Brooks, L. R. (1967). The suppression of visualization by reading. *Quarterly Journal of Experimental Psychology*, **19**, 289-299.

Brooks, L. R. (1968). Spatial and verbal components of the act of recall. *Canadian Journal of Experimental Psychology*, **22**, 349-368.

Cuddy, L. L., Cohen, A. J., & Mewhort, D. J. K. (1981). Perception of structure in short melodic sequences. *Journal of Experimental Psychology: Human Perception & Performance*, **7**, 869-883.

DeWitt, L. A., & Crowder, R. G. (1986). Recognition of novel melodies after brief delays. *Music Perception*, **3**, 259-274.

Dowling, W. J. (1978). Scale and contour: Two components of a theory of memory for melodies. *Psychological Review*, **85**, 341-354.

Dowling, W. J. (1984). Musical experience and tonal scales in the recognition of octave-scrambled melodies. *Psychomusicology*, **4**, 13-32.

Dowling, W. J. (1991a). Pitch structure. In P. Howell, R. West, & I. Cross (Eds.), *Representing musical structure* (pp. 33-57). London: Academic Press.

Dowling, W. J. (1991b). Tonal strength and melody recognition after long and short delays. *Perception & Psychophysics*, **50**, 305-313.

Dowling, W. J., & Bartlett, J. C. (1981). The importance of interval information in long-term memory for melodies. *Psychomusicology*, **1**, 30-49.

Dowling, W. J., & Fujitani, D. S. (1971). Contour, interval, and pitch recognition in memory for melodies. *Journal of the Acoustical Society of America*, **49**, 524-531.

Dowling, W. J., Lung, K. M.-T., & Herrbold, S. (1987). Aiming attention in pitch and time in the perception of interleaved melodies. *Perception & Psychophysics*, **41**, 642-656.

Edworthy, J. (1985). Melodic contour and musical structure. In P. Howell, I. Cross, & R. West (Eds.), *Musical structure and cognition* (pp. 169-188). London: Academic Press.

Koffka, K. (1935). *Principles of Gestalt psychology*. New York: Harcourt, Brace.

Krumhansl, C. L. (1990). *Cognitive foundations of musical pitch*. Oxford: Oxford University Press.

Macmillan, N., & Creelman, D. (1991). *Detection theory: A user's guide*. Cambridge: Cambridge University Press.

McColl, E., & Seeger, P. (1977). *Travellers' songs from England and Scotland*. Knoxville: University of Tennessee Press.

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, **63**, 81-97.

Monahan, C. B., & Carterette, E. C. (1985). Pitch and duration as determinants of musical space. *Music Perception*, **3**, 1-32.

Pechstedt, P. H., Kershner, J., & Kinsbourne, M. (1989). Musical training improves processing of tonality in the left hemisphere. *Music Perception*, **6**, 275-298.

Schneider, W., & Shiffrin, R. M. (1977). Controlled and automatic human information processing: I. Detection, search, and attention. *Psychological Review*, **84**, 1-66.

Shepard, R. N., & Teghtsoonian, M. (1961). Retention of information under conditions approaching a steady state. *Journal of Experimental Psychology*, **62**, 302-309.

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

Tovey, D. F. (1956). *The forms of music*. Cleveland: Meridian Books.

Trehub, S. E., Morongiello, B. A., & Thorpe, L. A. (1985). Children's perception of familiar melodies: The role of intervals. *Psychomusicology*, **5**, 39-48.

NOTES

1. Some might object that the stimuli in the present experiments are too simple to be called "melodies," since they are brief and isochronous. However, the musicologist Tovey (1956, p. 91), in distinguishing melodic from rhythmic aspects of music, defines melody as "the organization of successive musical sounds in respect of pitch." The present stimuli definitely meet that criterion. Furthermore, the seven-note melodies we use are not lacking in rhythmic structure, both in being accented on the first, fourth, and seventh notes and in carrying the inevitable accents that arise from turns in the pitch contour. It is also true that many "real" melodies have isochronous first phrases: "Jingle Bells" ("Dashing through the snow," 5 notes); "Now Thank We All Our God" (6 notes); "Twinkle, Twinkle," "Good King Wenceslaus," "Mary Had a Little Lamb," "London Bridge," "Old Macdonald," and "Aura Lee" ("Love Me Tender," all with 7 notes); "Little Moses" (8 notes); "She'll Be Comin' 'Round the Mountain" (11 notes); and "God Rest Ye Merry Gentlemen" (14 notes). Among numerous isochronous classical melodies are the opening theme of Beethoven's *Fourth Piano Concerto* (12 notes) and the fugue theme of Bach's *Toccata and Fugue in D Minor* (16 notes, omitting the interleaved notes, which included would make the total 31). Therefore, because the present stimuli meet a well-considered definition of melody, and because they are highly similar to patterns commonly referred to as melodies, it is preferable to call them melodies than to use some awkward circumlocution such as "isochronous tone sequences varying in pitch that closely resemble the first phrases of common melodies."

2. We conducted a pilot study just like Experiment 3, except that there was just one session containing a mixture of both tonal and atonal stimuli. That study showed a main effect of tonality (with better performance for tonal melodies than atonal), but no tonality × comparison type interaction. In all other respects, the results closely resembled those of Experiment 3. However, the data were rather noisy due to the low number of trials contributing to each data point.

(Manuscript received March 8, 1993;
revision accepted for publication August 23, 1994.)