

Tonal strength and melody recognition after long and short delays

W. JAY DOWLING

University of Texas at Dallas, Richardson, Texas

In a continuous-running-memory task, subjects heard novel seven-note melodies that were tested after delays of 11 sec (empty) or 39 sec (filled). Test items were transposed to new pitch levels (to moderately distant keys in the musical sense) and included exact transpositions (targets), same-contour lures with altered pitch intervals, and new-contour lures. Melodies differed in tonal strength (degree of conformity to a musical key) and were tonally strong, tonally weak, or atonal. False alarms to same-contour lures decreased over the longer delay period, but only for tonal stimuli. In agreement with previous studies, discrimination of detailed changes in pitch intervals improved with increased delay, whereas discrimination of more global contour information declined, again only for tonal stimuli. These results suggest that poor short-delay performance in rejecting same-contour lures arises from confusion that is based on the similarity of tonality between standard stimuli and lures. If a test item has the same contour and a similar tonality to a just-presented item, subjects tend to accept it. After a delay filled with melodies in other tonalities, the salience of key information recedes, and subjects base their judgments on more detailed pattern information (namely, exact pitch intervals). The fact that tonality affects judgments of melodic contour indicates that contour is not an entirely separable feature of melodies but rather that a melody with its contour constitutes an integrated perceptual whole.

Listeners are often confused in melody recognition between transpositions of novel melodies and lures that copy the contour (the ups and downs) of the target but not its interval sizes. For example, immediately after hearing the melody in Figure 1A, listeners tend to judge the melody in Figure 1B to be an exact transposition of Figure 1A, even though two pitches (and three intervals) have been altered in Figure 1B from those of an exact transposition (as indicated by the bracket). This confusion is particularly strong when the transposition is to a closely related key in the musical sense; that is, when the musical scale underlying the transposition is very similar to the scale underlying the original melody (Bartlett & Dowling, 1980). The confusion of melodies that are different in intervallic detail but are similar in contour and scale suggests that contour and scale are important features in memory for melodies (Dowling, 1978).

Figure 2 elaborates on the notion of scale and contour as important features of the pitch material of melodies. In the figure, parallel sets of features go from local characteristics of single melodies (in this case, the theme from Bach's "Little" G-minor Fugue, BWV 578) to global invariants that hold across sets of melodies, for example, the set of transpositions and contour-preserving imitations

of the fugue theme found in Bach's piece. (An "imitation" preserves pitch contour but not the exact interval sizes of a transposition.) The left side of Figure 2 shows successive abstractions from the interval pattern of Bach's melody: the contour (the pattern of ups and downs), the slope of the contour (the succession of melodic trajectories), and the set of contour inflections (or changes in slope). The right side of Figure 2 shows successive abstractions from the pitch pattern of Bach's melody: the key (the hierarchically ordered set of pitches), the tuning system (the unordered pitch set of the key) with its associated modes (or alternate hierarchical orderings), and the tonal scheme (the abstracted pitch pattern of the tuning system, which is not tied to any one pitch class). As Deutsch (1969) suggested, there is evidence for the cognitive relevance and separateness of each of these parallel schemes of invariants based on pitch and interval. As an example of the effects of features toward the bottom of the two columns, Cuddy, Cohen, and Mewhort (1981) demonstrated that both contour complexity (in the sense of number of contour inflections) and harmonic complexity (in the sense of key-region relationships within a melody) were important determinants of transposition recognition.

The confusion of transpositions with same-contour lures arises from the listener's reliance on contour and scale features and lies at the heart of two effects in recognition memory for melodies: the key-distance effect (mentioned above) and the shift of importance of contour and intervals in recognition with the passage of time. Key distance refers to the similarity of scales underlying a pair of melo-

I thank Se-Yeul Kwak for assistance in generating the stimuli and Melinda Andrews and James C. Bartlett for helpful comments on the design and the manuscript. Requests for reprints should be addressed to W. Jay Dowling, Program in Applied Cognition and Neuroscience, University of Texas at Dallas, Richardson, TX 75083-0688.

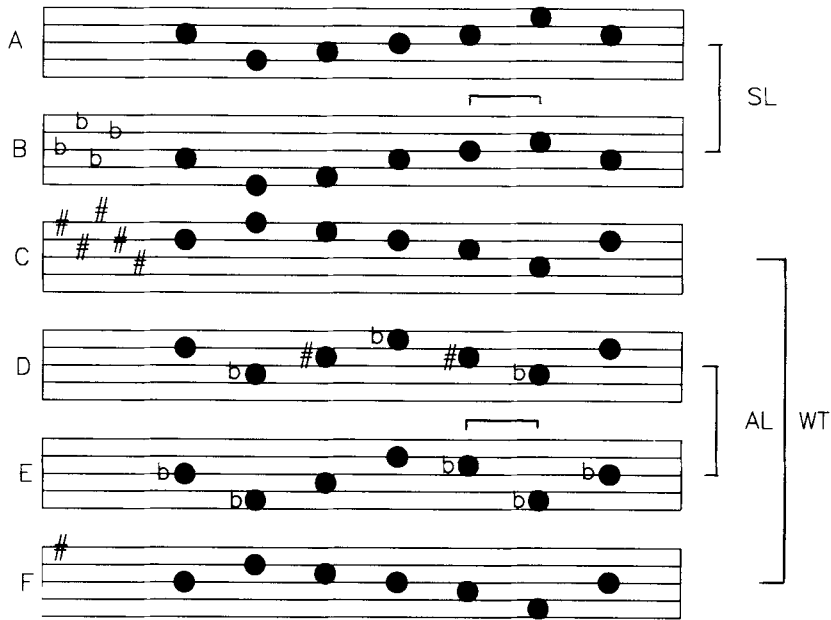


Figure 1. Stimulus patterns from the experiment arranged in the order in which they appeared in one of the lists. The bracket SL indicates a strongly tonal target tested with a same-contour lure, the bracket AL indicates an atonal target tested with a same-contour lure, and the bracket WT indicates a weakly tonal target tested with a target. (A) = the introduction of a novel strongly tonal melody; (B) = a short-delay test of Melody A with a same-contour lure (the bracket indicates the two altered pitches); (C) = the introduction of a novel weakly tonal melody; (D) = the introduction of a novel atonal melody; (E) = a short-delay test of Melody D with a same-contour lure; and (F) = a long-delay test of Melody D with a target (exact transposition). (A treble clef is understood.)

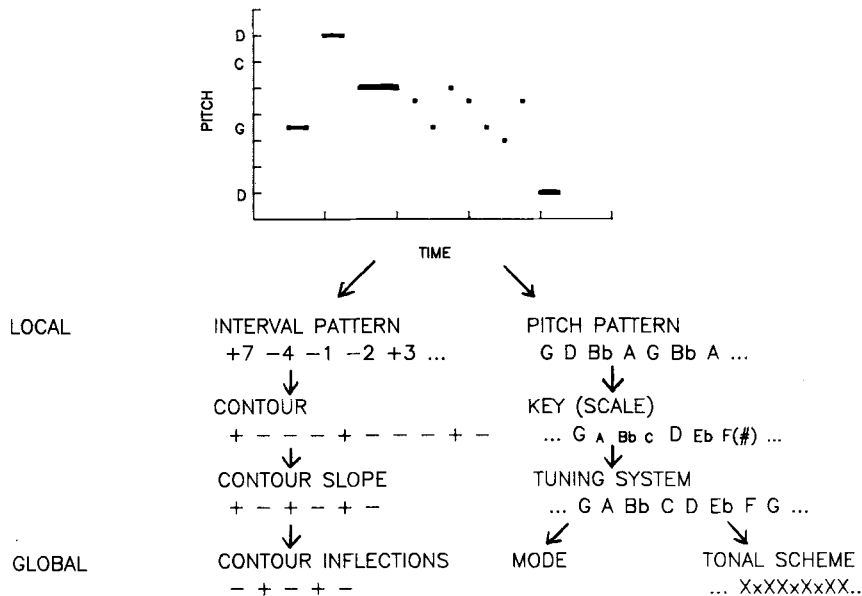


Figure 2. Schematic organization of the pitch material in melodies, going from local stimulus features of particular melodies at the top of the diagram to global invariants that hold across sets of melodies at the bottom. Invariants involving interval patterns are on the left; invariants involving pitch patterns are on the right. The melody is the theme from J. S. Bach's "Little" Fugue in G minor, BWV 578.

dies. The scales of closely related keys share relatively many pitches, and the scales of distant keys relatively few (for additional complexities, see Lerdahl, 1988, and Krumhansl, 1990a). Bartlett and Dowling (1980) tested novel melodies after short delays with transposed comparisons that were either in keys closely related to the standard or in distant keys. False-alarm rates to same-contour-different-interval lures decreased with increasing key distance, whereas hit rates to targets were unaffected by key distance. Although near- and far-key targets were equally recognizable, key similarity of lures led to confusion between lures and targets.

The second effect that depends on confusion in short-term recognition is the decline over time of the importance of contour (compared with exact pitch intervals). Dowling and Bartlett (1981) found that contour recognition, quite good when tested immediately, declined markedly after about 30 sec and that interval recognition after that long a delay was relatively good. Following a delay, listeners even had difficulty recognizing same-contour lures when they were instructed to do so. In a better controlled study, DeWitt and Crowder (1986) found even stronger evidence for the transition from domination of recognition by contour after short delays to relatively better memory for intervals after long delays.

The key-distance effect and the temporal shift away from domination of recognition by contour both depend on confusion arising from the listener's reliance on the relatively global features of scale and contour. Factors that dispel that confusion lessen or eliminate those effects. The confusion appears strongest with brief (5- to 10-note), novel melodies tested immediately after presentation. The confusion does not occur with highly familiar melodies (Dowling & Fujitani, 1971). Even methods in which the target melody is presented more than once (as with a standard followed by a forced choice between two comparisons) fail to obtain a key-distance effect as a result of this confusion (Cuddy et al., 1981; Cuddy, Cohen, & Miller, 1979). Confusion arising from contour similarity also tends to disappear with longer melodies in which the contour naturally becomes difficult to encode (Edworthy, 1985). Confusion as a result of scale similarity disappears when time and intervening context reduces the salience of the scale of the standard, which is essentially the effect obtained by Dowling and Bartlett (1981) and by DeWitt and Crowder (1986).

The present experiment was designed to test a further consequence of this view. If the shift between short-term and long-term recognition from contour to intervals is caused by confusion arising from scale similarity in the short-term case, then the effect should be attenuated by any factor that lessens that confusion. The salience of the musical scale should be greatest for melodies strongly representative of the tonality and less so for weakly tonal melodies. The scale should not be salient at all for melodies with pitch patterns that fall outside any consistent tonality (here called *atonal* melodies). As the salience of the scale decreases, confusions based on scale similarity

should decrease also. As tonal strength decreases, false-alarm rates to same-contour lures should decrease, too.

There is currently a debate in the psychological literature that complicates the task of defining "tonal strength." I incline toward a definition that sees tonally strong melodies as prototypical exemplars of a tonality. In this view, melodies that clearly establish a tonality by emphasizing its most important pitches and by traversing a standard cadence pattern in the key are tonally strong. Such melodies emphasize focal anchor pitches in the scale and the resolution of unstable pitches (e.g., the seventh scale degree, *ti*) to stable ones (the tonic, *do*). I believe this definition is in general agreement with the approaches of Cuddy et al. (1981), Krumhansl (1990a, 1990b), and Lerdahl (1988). Cuddy et al. used five levels of tonal strength, ranging from tonally strong seven-note sequences that began and ended on the tonic (*do*) or fifth (*sol*) and outlined principle triads in the key, such as tonic (I) and dominant (V), to atonal sequences that could not be easily assigned to any key. Cuddy et al. obtained corroboration for their a priori constructions in the analyses of three music theorists, as well as in the ratings of moderately and highly trained undergraduates. Cuddy et al. found that tonal strength thus defined affected listeners' ability to detect alterations in pitch intervals, ranging from around 90% correct for the tonally strong melodies to around 60% for the atonal ones (where chance was 50%).

An alternative approach is to emphasize rare, as contrasted with typical, intervals in a key as most essential to defining a tonality. Butler and Brown (1984; Brown, 1988; Butler, 1989) emphasize the fact that in an appropriate harmonic context, rare intervals (rare in the sense that they occur relatively rarely in European tonal music) such as the tritone (6 semitones) can point unambiguously to the tonal center. My view is that in this formulation, the appropriateness of context is very important (as Butler, 1990, p. 330, emphasizes) and that what sets up the context is the preponderance of relatively predictable elements. Actual music is predominantly predictable, with a sprinkling of rare pattern elements. I believe the best definition of a tonally strong pattern is one that clearly represents the predictable, prototypical aspects of a tonality.

A third method for varying tonal strength that resembles the first method more than the second is that used by Watkins (1985, Experiment 2). Watkins restricted the tonally stronger ("low fifth-span") melodies to limited key regions with respect to the circle of fifths. That is, tonally stronger melodies remained close to the region of a single key. Tonally weaker ("high fifth-span") melodies wandered beyond the boundaries of a single key region. Watkins found that as long as standard tuning was preserved, performance was best with the tonally stronger melodies.

Though inclining toward the definition of Cuddy et al., I found it extremely difficult to duplicate the harmonic outlines they used and simultaneously use 60 different contours in writing plausible, tonally strong melodies. Since

I believed it important to avoid duplicating contours within the experiment, what I did was to write tonally strong items that started and ended on the tonic, that constituted clearly tonal patterns in the key, and that were “melodious” in the sense of using relatively small pitch intervals (Pechstedt, Kershner, & Kinsbourne, 1989). The tonally strong patterns were then weakened to form tonally weak and atonal patterns. Tonally weak patterns were formed by shifting the tonally strong patterns along the diatonic scale so as to begin and end on the third scale degree (*mi*) instead of the tonic (in effect changing their mode from the common major mode to the uncommon phrygian). Atonal patterns were formed by replacing the tonal scale of the tonally strong patterns with a pseudoscale whose intervals could not occur in any consistent tonality.

Tonally strong and atonal stimuli here are similar to Watkins’s (1985, Experiment 1) “diatonic” and “non-diatonic” stimuli. Watkins found much better performance with diatonic melodies, as long as standard interval sizes for half steps were preserved. If tuning was expanded or shrunk, performance with the two types of stimuli (now neither of them was congruent with the standard diatonic scale) was about the same.

Figure 1 shows examples of trials from the present experiment. In A, we see a tonally strong novel melody tested immediately in B by means of a lure with the same contour but different intervals. (The two pitches marked by the bracket in B are altered from what they would have been in an exact transposition.) C shows a weakly tonal novel melody that is tested after a filled delay by an exact transposition in F. The pair D-E (with yet another contour) involves an immediate test of an atonal melody. The subjects’ recognition performance was measured by their ability to say “old” to genuine transpositions (such as F) and “new” to lures. Here B and E functioned as same-contour-different-interval lures. C and D served as new-contour lures for earlier trials involving weakly tonal and atonal items, respectively.

If the failure of the subjects to discriminate transpositions from same-contour lures after short delays was caused by confusion produced by similarity of key between standard and comparison melodies, then, as tonal strength was weakened, key similarity should have less effect, and less confusion would result. The extreme case would involve atonal melodies that could not be interpreted in any single key. With longer delays, the confusion arising with tonal melodies should disappear, and performance should improve as a result of the decline of false-alarm rates to same-contour lures.

Croonen and Kop (1989) made a contrasting claim regarding the effects of tonal strength on memory for the interval patterns of melodies—namely, that subjects would encode intervals in tonally strong melodies rapidly and accurately—and therefore, the confusion observed by Bartlett and Dowling (1980; Dowling & Bartlett, 1981) should not appear at short test delays with strongly tonal stimuli. Unfortunately, the following problems in Croonen and Kop’s experimental design prevented their study from serving as an adequate test of that proposition:

1. As noted above, pitch intervals in familiar melodies are remembered much better than those of novel melodies heard for the first time. However, Croonen and Kop used the same two melodies repeatedly throughout their experiments, thereby familiarizing subjects with the test materials and obviating a test of the hypothesis.

2. Croonen and Kop’s (1989, pp. 58 and 63) response to such anticipated criticism was that the number of false alarms to same-contour-different-interval lures did not decrease with trials, as would have been expected on the supposition that increasing familiarity leads to better interval learning. Unfortunately, they used only one order of trials, so order was confounded with particular items in the set of lures. Since they used same-key-different-mode imitations constructed by shifting the origin of the melody along the scale to each of the six other scale degrees, some of the lures were undoubtedly more strongly tonal than others, and order was thus confounded with tonal strength.

3. The use of only two standard melodies, both tonally strong, makes generalization to other melodies problematic and obviates conclusions concerning the comparative effects of tonal strength.

For these and other reasons, Croonen and Kop’s study does not provide a test of the relevant issues. Their useful suggestion that tonal strength may be an important variable affecting the time course of memory for melodic intervals remains to be evaluated, which is a major purpose of the present experiment.

The present experiment was designed to measure memory for melodic contour and melodic interval sizes following both long and short delays. Short delays consisted of unfilled time intervals of 11.0 sec; long delays consisted of filled intervals averaging 39.0 sec. Memory for contour was assessed by comparing correct recognitions of transpositions (such as Melody F in Figure 1) with false alarms to different-contour lures of the same tonal strength (such as C in Figure 1). Memory for intervals was assessed by comparing correct recognitions of transpositions with false alarms to same-contour-different-interval lures (such as B and E) of the same tonal strength as the targets. Same-contour lures were formed by altering the pitches of the fourth and fifth notes of the melody from what they would have been in a transposition (brackets in Figure 1, B and E).

The trials were arranged in a continuous-running-memory task (Shepard & Teghtsoonian, 1961), as illustrated in Figure 3. There, brackets connect standard stimuli and comparisons for each test. Short-delay pairs

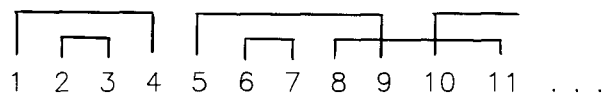


Figure 3. Organization of part of the list in the continuous-running-memory task. Trials 1, 2, 5, 6, 8, and 10 introduce new-contour melodies; Trials 3 and 7 are short-delay tests of melodies introduced on Trials 2 and 6; and Trials 4, 9, and 11 are long-delay tests of melodies introduced on Trials 1, 5, and 8.

had no intervening items. Long-delay pairs were separated by an average of two intervening items, and the list was irregularly arranged so the subjects could not anticipate the nature of each trial. The subjects responded on each trial, and the responses to first members of pairs provided false-alarm rates to new-contour lures. (No contour was repeated within the 122 trials of the experiment.)

There are several advantages to continuous running memory as a method. Since the subjects could not know which trials contained standards and which trials contained comparisons, their perceptual set was the same for each melody, and they could not use special strategies on "test" trials, as contrasted with "input" trials, thereby meeting a valid objection of DeWitt and Crowder (1986) to Dowling and Bartlett's (1981) "inside/outside" method. Furthermore, the subjects were occupied with responding on each trial, which interfered with the implementation of more than rudimentary "rehearsal" strategies. The subjects also found the method very congenial, since it moved along quickly and lacked the stress of a particular set of test trials.

On the basis of the above arguments, I expected that test delay and type of lure would interact, so that contour recognition would decline and interval recognition would improve with increasing delay, replicating the results of Dowling and Bartlett (1981) and DeWitt and Crowder (1986). The improvement should be mostly due to a decline of false alarms to same-contour lures over time. Also, on the supposition that poor short-term performance in interval discrimination arises from confusion caused by similarity of tonality, the change in same-contour false-alarm rates over time should be especially true of strongly tonal items. Stimuli with lower tonal strength should have less similarity of tonality and thus should generate less confusion. Finally, the weaker the tonality, the less well remembered should be the melody, which agrees with Cuddy et al. (1981) and Watkins (1985). Even contour should be less well remembered for the atonal melodies, in keeping with Bartlett and Dowling's (1988) result that contour changes were more difficult to recognize in atonal (vs. tonal) test melodies.

METHOD

The experiment was arranged in a 2 experience levels \times 2 test delay (short, long) \times 3 tonality conditions (strong, weak, atonal) \times 2 test comparisons (target vs. new, target vs. same-contour lure) design. All but the first of the variables involved within-groups comparisons. The 120 trials of the continuous-running-memory task consisted of 60 intermingled pairs of items, in which each pair contained an initial item with a novel contour and a test item that was either an exact transposition of the first or a contour-preserving lure (a transposition with two altered pitches). Positive responses to exact transpositions provided hit rates to targets, and positive responses to contour-preserving lures provided false-alarm rates for the target-versus-same-contour comparison. False-alarm rates for the target-versus-new-contour comparison were provided by positive responses to new-contour initial members of pairs, categorized according to tonality level and temporal separation from preceding items of the same tonality condition.

Subjects

Eighty-four undergraduates at the University of Texas at Dallas (mean age 29.5 years) served as subjects to fulfill part of their course requirements in psychology. Forty-five had more than 2.0 years of explicit musical training (mean = 7.0 years, SD = 4.3 years) and were categorized as musically experienced. The remaining 39 subjects were categorized as musically inexperienced. The subjects served in group sessions. (Note that the present definition of "inexperienced" is roughly the same as Croonen and Kop's, 1989, definition of "moderately experienced.")

Stimuli

Each stimulus consisted of a seven-note melody presented at a rate of 3.0 notes/sec (see Figure 1). The final note of each melody was twice the duration of the first six notes. Such melodies group themselves rhythmically as two triplets followed by a longer note. A silent response interval of 11.0 sec followed each stimulus in the list. Stimuli were produced by a Commodore 64 computer producing triangular waves via its 6581 sound-interface device. The stimuli were recorded on tape and presented to the subjects via loud-speaker at comfortable levels.

The stimuli were presented in a continuous-running-memory task (Shepard & Teghtsoonian, 1961) of 120 trials, preceded by four sample stimuli illustrating the contrasts shown in Figure 1, and 2 buffer trials containing novel melodies that were not scored. The 120 valid trials consisted of 60 pairs of melodies such that the first member of each pair introduced a novel melody that was different in contour from all of the 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 and two arbitrarily selected contours that had just one contour inflection. The second member of the pair tested memory for the first member and shared its contour. The test stimulus was either an exact transposition of the initial melody or was a transposition with the fourth and fifth pitches altered by 1.0 to 4.0 semitones. In the continuous-running-memory list, each initial member of a pair was followed by its mate either immediately or after a delay of (usually) two intervening items (a filled interval of 39.0 sec).

The melodies in each pair were either strongly tonal, weakly tonal, or atonal. Strongly tonal melodies conformed to the diatonic major scale and began and ended on the first scale degree (*do*, the tonic). Weakly tonal melodies also conformed to the major scale but began and ended on the third scale degree (*mi*). Atonal melodies began and ended with the same pitch class but were based on a set of pitches that guaranteed that none conformed to any single tonality. The pitches of that set were arranged either 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 \flat , E, F, G \flat , A, B. The two clusters with 1-semitone spacing (B-C-D \flat and E-F-G \flat) ensure that the resultant pattern cannot conform to a tonal scale.

In generating the 60 pairs of melodies, the experimenter first constructed a melody pair for each contour, and an attempt was made to achieve melodies that would be tonally strong in the sense meant by Cuddy et al. (1981), especially by using pitches from the principal triads in the key in a coherent order and by starting and ending 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 et al., 1989) in the sense of having relatively narrow pitch intervals and forming relatively coherent gestalts. The second member of the pair was formed by altering the pitch of the fourth and fifth notes of the first melody. Both were designed to be equally strong, but as a safeguard against possible asymmetries of strength, half of the subjects heard the melodies in reverse order in the list (see counterbalancing scheme below).

Twenty melody pairs were randomly assigned to each of the three tonality conditions. Tonally weak melodies were formed by shifting the initially generated tonally strong melodies along the diatonic scale so as to start on the third degree of the scale and to preserve diatonic (but not semitone) pitch intervals. Atonal melodies were formed by altering their pitch patterns to conform to the atonal "scale" pattern described above.

The 20 melody pairs of each tonal strength were randomly divided into two groups of 10 to be tested as targets and as same-contour lures. Each of those groups of 10 was again divided randomly, and each half was assigned to short-delay and long-delay tests. Thus, the assessments of target hit rates and same-contour false-alarm rates for each subject were each based on five responses. 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 the 12 melody-pair types defined by the dimensions of tonality, comparison type, and test delay.

There were six long-delay pairs in each block: four contained two intervening items (a 39.0-sec delay), one contained one intervening item (25.0 sec), and one contained three intervening items (53.0 sec). Over the 120 trials, each of the six long-delay pairs in a tonality \times comparison-type condition was tested with a mean delay of approximately 39.0 sec (range = 33.4–41.8 sec).

New-contour lures were divided within each tonality condition into items that immediately followed an item of the same tonality and into items that followed items of the same tonality by delays of one to six intervening items. The former constituted the set of short-delay new-contour lures; the latter constituted the set of long-delay new-contour lures. New items not meeting the constraints were discarded. There were 4–11 members of each tonality \times delay set for each list.

All comparison members of tonal pairs were in keys moderately distant from their initial members, with a distance of three or four steps around the circle of fifths from the origin. Thus, comparison melodies always started and ended on a pitch 3 or 4 semitones above or below that of the initial melody of the pair. Furthermore, successive trials were in keys (or based on starting notes of atonal melodies) that were 3 or 4 semitones apart (and their keys were three or four steps apart around the circle of fifths). Each of the 12 possible keys appeared between 9 and 11 times in the series of 120 trials.

For purposes of counterbalancing, different groups received different versions of the list. Approximately half of the subjects received each list in reverse order, so that comparison melodies became initial members of pairs and vice versa. And for approximately half of the subjects, the membership of short- and long-delay pairs was reversed, so that effects attributed to test delay could not be due to the memorability of particular items. Four groups of 18–26 subjects (about equally divided between inexperienced and experienced) received each of the four types of list thus generated.

Procedure

Subjects were introduced to the experiment by brief explanations of the continuous-running-memory task, the difference between targets and same-contour lures, and the confidence-level response scale. The experimenter explained the intermingling of new items and test items by using a diagram similar to Figure 3. He informed the subjects that as many as "four or five" items might intervene before an item was tested, but that the interval would never be longer than that. By means of the familiar tune "Twinkle, Twinkle, Little Star," the experimenter demonstrated the difference between the exact transposition of a melody and a same-contour 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). Unlike the subjects in Dowling and Bartlett (1981), the present subjects were instructed to respond "old" only to exact transpositions of melodies that they heard previously in the list. The experimenter explained the use of the 6-

category confidence-level scale with the responses "Very sure old," "Sure old," "Old," "New," "Sure new," and "Very sure new." After the instructions, the experimenter introduced four sample trials that illustrated the comparison types of exact transposition, same-contour lure, and different contour. The samples were strongly tonal items with contours that were not used in the rest of the experiment. The subjects responded with response-category numbers on a numbered answer sheet. To keep everyone synchronized, the experimenter announced the trial number about every 25 trials.

The subjects performed the 122 trials of the experiment (2 buffers at the beginning and 120 valid trials) in 28.5 min and also completed a brief questionnaire concerning musical experience. On the questionnaire, the subjects were asked to specify the ages at which they had begun their musical training and performance.

Data Analysis

I analyzed the data in two ways. First, I examined confidence-level rating data for the two types of lure (same-contour and new-contour) for the three tonality levels and the two delay conditions by means of an analysis of variance (ANOVA). Since a principal issue was the presence of tonality-induced confusion leading to the acceptance of same-contour lures in the short-delay condition, I planned comparisons of short- versus long-delay same-contour lures for the three levels of tonality, as well as parallel comparisons for the new-contour lures.

Second, to assess the accuracy of memory performance, I used area under the memory operating characteristic (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 of the 3 tonality \times 2 delay trial types, giving 12 area scores per subject. One area compared hit rates to targets and false-alarm rates to same-contour lures; the other compared hit rates to targets and false-alarm rates to new-contour lures in the same tonal-strength condition. Areas under the MOC were evaluated by means of an ANOVA. Because of the theoretical issues involved, I planned comparisons of accuracy in distinguishing targets from same-contour lures in short- versus long-delay conditions for each level of tonal strength.

RESULTS

Ratings of Lures

The most direct test of the hypothesis that lower performance on melodic-interval recognition at short delays was brought about by tonality-induced confusion was to look at responses to same- and new-contour lures. Ratings of lures on the 6-point confidence-level scale were analyzed by means of a 2 experience levels \times 3 tonal-strength levels \times 2 test delays \times 2 lure types ANOVA. Experience involved a between-groups comparison, whereas the other variables involved within-groups comparisons. Main effects of tonal strength [$F(2,164) = 21.25, p < .0001$] and lure type [$F(1,82) = 79.39, p < .0001$] were significant. Weakly tonal lures generated higher ratings (mean = 3.59) than either strongly tonal (3.21) or atonal (3.31) lures; same-contour lures had higher ratings than new-contour lures (3.52 vs. 3.22).

The interaction of lure type \times delay was significant [$F(1,82) = 26.78, p < .0001$], and the mean ratings for targets, as well as for lures, are shown in Table 1. In accordance with the tonality-confusion hypothesis, ratings of same-contour lures declined from short to long delays, so that ratings for the two types of lures converged over

Table 1
Ratings on a 6-Point Confidence-Level Scale for Targets and Same- and New-Contour Lures at Two Delays of Test

Test Delay	Targets	Lures	
		Same Contour	New Contour
Short	3.99	3.67	3.14
Long	3.95	3.37	3.31
Mean	3.97	3.52	3.22

Table 2
Ratings on a 6-Point Confidence-Level Scale for Targets and Same- and New-Contour Lures for Subjects at Two Levels of Experience

Experience	Targets	Lures	
		Same Contour	New Contour
None	3.98	3.51	3.32
> 2 years	3.96	3.53	3.14
Mean	3.97	3.52	3.22

Table 3
Areas Under the MOC for Targets versus New- and Same-Contour Lures for Short- and Long-Delay Tests

Delay	Target vs. New Contour	Target vs. Same Contour	Mean
Short	.65	.56	.61
Long	.63	.59	.61
Mean	.64	.58	

the delay period. This effect is broken down by tonality condition in Figure 4, in which it is clear that the decrease in same-contour lures occurred only with tonal melodies. Planned comparisons of mean ratings of lures within each tonality level disclosed that the differences between ratings of same-contour lures at the two delays were significant only for tonally strong [$t(83) = 3.40, p < .01$] and tonally weak [$t(83) = 2.27, p < .05$] melodies and not for atonal ones [$t(83) = 1.78, n.s.$]. None of the differences was significant for new-contour lures.

The only remaining significant effect was the interaction of lure type \times experience [$F(1,82) = 9.98, p < .01$], which is shown in Table 2. The experienced subjects were better able than the inexperienced subjects to reject new-contour lures.

A similar ANOVA that included ratings of both targets and of lures disclosed all of the significant effects mentioned above, with the sole addition of an interaction of item type \times tonal strength [$F(4,328) = 16.02, p < .0001$] that was brought about by the fact that, for targets, ratings declined with declining tonal strength (mean ratings of 4.17, 3.92, and 3.81 from strong to atonal), in contrast with the pattern described above for lures (Figure 4).

Area Under MOC

The areas under the MOC were analyzed by means of a 2 experience levels \times 3 tonal-strength levels \times 2 test delays \times 2 test comparisons ANOVA. The main effects

of tonality [$F(2,164) = 19.17, p < .0001$] and test comparison [$F(1,82) = 105.48, p < .0001$], were significant. Tonally strong melodies were better recognized than tonally weak or atonal melodies (areas of .67, .56, and .60, respectively), and targets were easier to distinguish from new-contour lures (.64) than from same-contour lures (.58). No other main effects were significant.

The only significant effect involving experience was the interaction with test comparison [$F(1,82) = 10.47, p < .01$]. The experienced subjects performed better than the inexperienced subjects with new-contour comparisons (.66 vs. .63) but no better with same-contour comparisons (both areas = .58).

The interaction of delay \times test comparison was significant [$F(1,82) = 8.54, p < .01$], as shown in Table 3. The difference in performance levels for the two types

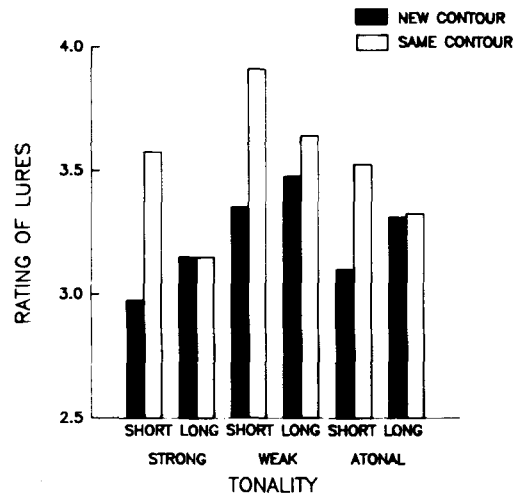


Figure 4. Ratings of new-contour (filled bars) and same-contour (open bars) lures after short and long delays for three levels of tonal strength.

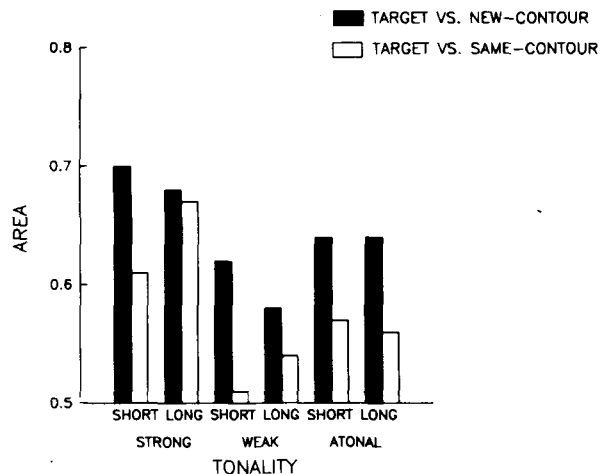


Figure 5. Areas under the MOC for targets having three levels of tonal strength tested after short and long delays against new-contour (filled bars) and same-contour (open bars) lures.

of comparisons was greater at the short delay (.09) than at the long delay (.04).

The only remaining significant effect was the interaction of tonal strength \times delay \times test comparison [$F(2,164) = 3.77, p < .025$] shown in Figure 5. In that interaction, the pattern, as shown in Table 3, of a lessening difference between the two target-lure comparisons over the delay period was true only for tonal melodies. Of the planned comparisons of performance across the delay interval, only that for the strongly tonal targets versus the same-contour lures was significant [$t(83) = 2.177, p < .05$]. That improvement in performance reflects the decrease with delay in false alarms to same-contour lures with increased delay, as noted above.

DISCUSSION

There are two main conclusions to be drawn from these results. First, false alarms to same-contour-different-interval lures decreased over the longer delay period. This decrease was reflected in the improvement in discrimination between targets and same-contour lures from shorter to longer delays. Second, tonal strength was an important determinant of this effect. In fact, it was for strongly tonal melodies that same-contour false alarms declined most dramatically (Figure 4), and only for strongly tonal melodies that the improvement in target-versus same-contour discrimination was statistically significant (Figure 5).

These results converge with other evidence (discussed above) that musical-scale information can cause confusion in short-term recognition-memory judgments. As the salience of key information fades into the background during a long delay filled with other melodies, accuracy in interval-recognition performance improves. As we would expect, the confusion due to musical key information was strongest for the strongly tonal melodies, precisely where key information itself was strongest and most salient. As expected, there was no such confusion with atonal melodies. Weakly tonal melodies displayed both of the negative effects of tonality. They were sufficiently tonal to invoke key-related short-delay confusion, but since they were only weakly tonal, their intervals were not well encoded. Hence, interval-recognition performance with them was distinctly inferior to that for strongly tonal items.

These results permit us to address an issue raised by DeWitt and Crowder (1986) in their search for an explanation of the puzzling phenomenon of the improvement with time of the difficult discrimination between transpositions and same-contour lures. DeWitt and Crowder noted that same-contour lures had typically been constructed by sliding the contour of the target along the scale to preserve diatonic steps but not exact intervals. In the process, mode was changed. ("Mode" refers to the positioning of the principal notes of the melody with respect to the interval pattern of an underlying tuning system; see Dowling, 1978.) DeWitt and Crowder suggested that since transpositions preserved mode and this type of same-

contour lure did not, their relative discriminability after a delay might simply mean that mode information persisted longer than contour information. The present experiment tested one aspect of this hypothesis by keeping mode constant between standard and comparison melodies, whether they were transpositions or same-contour lures. Here, same-contour lures were formed by altering two pitches in the middle of the melody, which was otherwise left unaltered (see Figure 1). Even with mode held constant, the results of Dowling and Bartlett (1981) and DeWitt and Crowder (1986) concerning the relative importance of contour and interval information after short and long delays were replicated here. Therefore, that effect does not depend on the effects of mode information on interval-based judgments of melodies.

As shown in Table 2, the experienced subjects performed better than the inexperienced subjects in rejecting new-contour lures but not in rejecting same-contour lures. It is perhaps surprising that where the more experienced subjects excelled was not in the subtle areas of recognizing transpositions and distinguishing them from same-contour lures but in using the more global contour information that is involved in rejecting new-contour lures. However, in a wide variety of melody-recognition tasks, more experienced subjects have had greater success using contour information than have less experienced subjects (Dowling, 1984). With musical training, subjects come to encode melodic intervals in terms of diatonic scale categories, and, at high levels of experience, this ability leads to enhanced performance with tonal materials (Cuddy et al., 1981; Dowling, 1986). But with the brief melodies of the present experiment, each remaining within a single key, the inexperienced subjects performed as well as the moderately experienced subjects in noticing changes in intervallic detail at both short and long delays. It may be that contour lends itself to explicit, perhaps verbal, encoding by subjects with musical training and that such encoding improves their memory for contour. It may also be, and the present data do not rule this out, that subjects with moderate amounts of musical training (such as the moderately experienced subjects of Dowling, 1986) may have been more susceptible than naive subjects to the confusions arising from tonality.

Finally, there is a broader conclusion that may be drawn from these results. Clearly, the melodies were heard as integrated whole patterns—gestalts—in which features such as contour and tonality interacted to affect recognition. Theories of perception and memory for melodies would have been much simpler if contour had been a separable feature in Garner's (1974) sense. In that case, we could have thought of the pitch material in a tonal melody as generated simply by the hanging of a contour on a tonal scale framework (Dowling, 1978) and of melody perception as a matter of simply encoding the contour and the mode (the relationship of contour and scale). However, it is clear from the present results, as well as from the results of Bartlett and Dowling (1988), that contour and tonality are not encoded and remembered in-

dependently. Bartlett and Dowling found that subjects had more difficulty noticing contour changes in atonal comparison melodies than in tonal ones. And in the present experiment, not only did hit rates to transpositions (in the form of confidence-level ratings) decline with weakening tonality, but false-alarm rates to new-contour lures rose. The lowest false-alarm rates to new-contour lures occurred with strongly tonal melodies, and the highest with weakly tonal melodies. As a result, proportion correct (in the form of area under the MOC) for same-versus-different-contour discrimination was best with strongly tonal melodies and worst with weakly tonal ones.

We might naturally expect tonality conditions to affect memory for intervals, but here we find tonality affecting memory for contour. The melodies were all constructed with approximately the same interval sizes, and their contours were randomly sampled from the same pool. Therefore, we must conclude that the tonal configuration of the melody—the degree to which the pitches of the melody match or mismatch the patterns of tonal invariants in the listener's experience—affects the way in which the contour of the melody is heard and encoded. It would be easy to imagine a listener who just encoded the contours of melodies ignoring the precise patterns of pitch intervals (as long as they were of moderate size). But that is not what happened. Both here and in Bartlett and Dowling's (1988) study, listeners did not ignore interval information in encoding contour, and the contours of melodies that matched tonal scale invariants of the culture were best remembered. This strongly suggests that melodies are heard as integrated wholes in which the various perceptual features interact.

REFERENCES

- 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., & DOWLING, W. J. (1988). Scale structure and similarity of melodies. *Music Perception*, *5*, 285-314.
- BROWN, H. (1988). The interplay of set content and temporal context in a functional theory of tonality perception. *Music Perception*, *5*, 219-250.
- BUTLER, D. (1989). Describing the perception of tonality in music: A critique of tonal hierarchy theory and a proposal for a theory of intervallic rivalry. *Music Perception*, *6*, 219-242.
- BUTLER, D. (1990). Response to Carol Krumhansl. *Music Perception*, *7*, 325-338.
- BUTLER, D., & BROWN, H. (1984). Tonal structure versus function: Studies of the recognition of harmonic motion. *Music Perception*, *2*, 5-24.
- CROONEN, W. L. M., & KOP, P. F. M. (1989). Tonality, tonal scheme, and contour in delayed recognition of tone sequences. *Music Perception*, *7*, 49-68.
- 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.
- CUDDY, L. L., COHEN, A. J., & MILLER, J. (1979). Melody recognition: The experimental application of musical rules. *Canadian Journal of Psychology*, *33*, 148-157.
- DEUTSCH, D. (1969). Music recognition. *Psychological Review*, *76*, 300-307.
- 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. (1986). Context effects on melody recognition: Scale-step versus interval representations. *Music Perception*, *3*, 281-296.
- 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.
- 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.
- GARNER, W. R. (1974). *The processing of information and structure*. New York: Wiley.
- KRUMHANSL, C. L. (1990a). *Cognitive foundations of musical pitch*. New York: Oxford University Press.
- KRUMHANSL, C. L. (1990b). Tonal hierarchies and rare intervals in music cognition. *Music Perception*, *7*, 309-324.
- LERDAHL, F. (1988). Tonal pitch space. *Music Perception*, *5*, 315-350.
- PECHSTEDT, P. H., KERSHNER, J., & KINSBOURNE, M. (1989). Musical training improves processing of tonality in the left hemisphere. *Music Perception*, *6*, 275-298.
- 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.
- WATKINS, A. J. (1985). Scale, key, and contour in the discrimination of tuned and mistuned approximations to melody. *Perception & Psychophysics*, *37*, 275-285.

(Manuscript received January 2, 1991;
revision accepted for publication May 20, 1991.)