Recognition of melodic transformations: Inversion, retrograde, and retrograde inversion*

W. J. DOWLING
University of California, Los Angeles, California 90024

The melodic transformations of inversion, retrograde, and retrograde inversion occur in pieces of music. An important question is whether such manipulations of melodic material are perceptually accessible to the listener. This study used a short-term recognition-memory paradigm and found that in the easier conditions all these transformations were recognized with better than chance accuracy. The ascending order of difficulty was: inversion, retrograde, retrograde inversion. There was no evidence that listeners distinguish between transforms that preserve the exact interval relationships of the standard stimulus and those that merely preserve its contour (pattern of ups and downs). In view of the order of difficulty of the transforms, two theoretical explanations of performance are possible (1) Listeners may perform the mental transformation required by the recognition task on a representation of the vector of pitches in the standard—an operation that is very like transforming a mental image of the written notation. (2) Listeners may handle inversions differently from the other transformations, comparing the standard and the comparison contour element by contour element, in temporal order. In this view, the temporal dimension would appear to have precedence over the pitch dimension in the musical structure, in consideration of the consequences of disturbing it.

How the perceiver recognizes the stimuli in his environment under their various transformations of shape and size is a basic problem in the study of perception. One version of this problem in audition is how the listener recognizes melodic transformations commonly used in music. In music, melodies already known to the listener may be speeded up or slowed down—diminution and augmentation. The pitch-interval sizes between the notes of the melody may be altered. Melody recognition with altered pitch relationships has been studied by Francès (1958), White (1960), and Dowling & Fujitani (1971). These transformations may operate on a whole melody or on only isolated elements—as in the temporal spacing of the intact phrases of a hymn in a baroque chorale prelude, or as in the alteration of only certain pitch relationships in the "tonal answer" to a fugue subject (Tovey, 1956). Another set of transformations operates on the melodic pattern as a whole, turning it upside down, backwards, or upside down backwards in the pitch-duration domain: the transformations of inversion, retrograde, and retrograde inversion. Of these, inversion is the most common and is found in nonwestern (for example, Indonesian) as well as in western music. Retrograde and retrograde inversion, though more rare, can be found throughout the history of western music. Tovey (1956) gives several examples from various periods. Figure 1 shows examples of these transformations such as were used in the present experiment: (C) is an inversion of (A), (E) is a retrograde of (A), and (G) is a retrograde inversion of (A).

One issue raised by the present study is whether transformations of melodic material can function as an actually perceived aspect of musical structure or whether they must be relegated to the category of merely intellectual conceits of the composer. The case is strongest for inversions. Dowling (1971) has shown that listeners do better than chance in recognizing inversions in a task similar to that of the present study. And the inversion transformation is very similar to the "mirror" transformation Restle (1970) includes in his structural theory of serial pattern learning. (The difference lies in the fact that Restle's theory allows the mirror transformation of a single element, whereas melodic inversion is only possible on a pattern of at least two elements.) There is more question concerning the perceptual accessibility of retrogrades and retrograde inversions. In various periods of western music, these transformations have figured in the construction of riddle canons and the like, where the composer wrote down only a cryptic set of notes with enough hints so that the whole piece could be reconstructed, as for example with Machaut's hint "Ma fin est mon commencement [Apel, 1944]." For

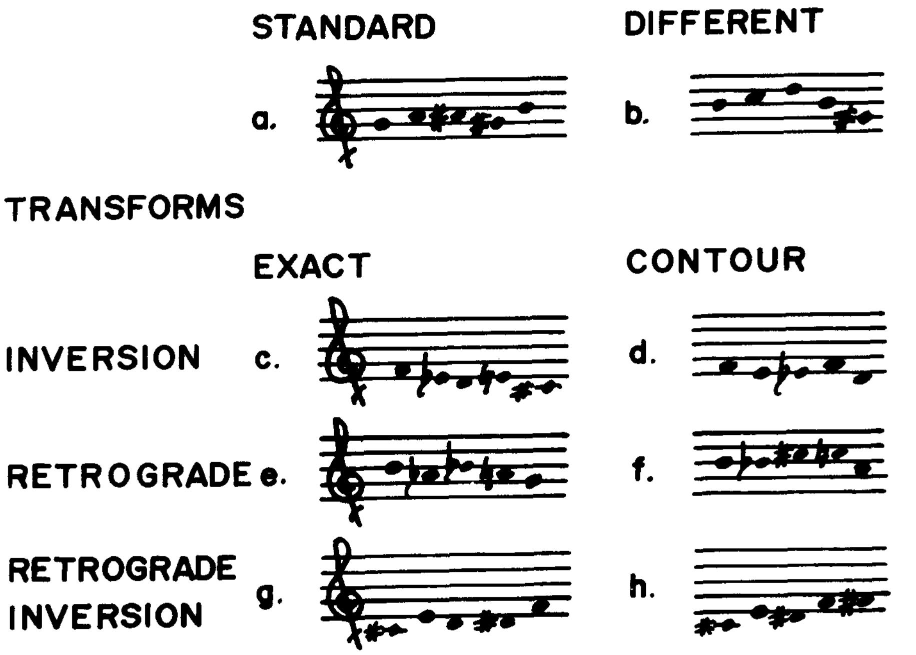

Fig. 1. Examples of stimuli used in the experiment. (A) Standard stimulus; (B) different comparison stimulus; (C), (E), and (G) exact-interval size-preserving transforms of the standard; (D), (F), and (H) contour-preserving transforms of the standard.

*This research was supported in part by a grant from the UCLA Academic Senate. Computing facilities were provided by the Department of Psychology, UCLA. I thank Michael Lisinsky, Edward Carterette, Caroline Dowling, Dan Rourke, James Thomas, Roberta Greer, Kelyn Roberts, and Dane Harwood for valuable assistance and comments.

examples of riddle canons by Bach, see David and Mendel (1966). Tovey (1956) doubts that retrogrades and retrograde inversions are effective perceptually, except in a very few cases, and in those cases he points out that rhythm seems to be the decisive factor, not melody. And Francès (1958) expresses reservations about the effectiveness of these melodic transformations as a cohesive force in 12-tone compositions. The present study will not attempt to demonstrate the effectiveness of these transformations in actual music, but only the existence of one of the necessary conditions of that effectiveness: the fact that listeners can recognize such transformations when they hear them under certain limited circumstances.

A second issue raised here is: given that listeners can succeed in recognizing these three types of transformation, what psychological processes might be involved? There are two convenient ways of characterizing melodic transformations in formal terms: as operations on intervals and as operations on pitches. The former is more convenient for purposes of formal description, but there is reason to believe that people tend to think and perceive in terms of the latter. The characterization of melodic transformations in terms of pitches can be derived by first assigning pitch numbers in semitones above some arbitrary reference tone, for example, middle C. Figure 1A would then be represented as:

$$7\ 9\ 10\ 8\ 11\ . \qquad (1)$$

The inversion of (1), shown in Fig. 1C, is obtained by subtracting the pitches of (1) from a constant, for example, 12:

$$5\ 3\ 2\ 4\ 1\ . \qquad (2)$$

The retrograde of (1) (Fig. 1E) is obtained by reversing the order of pitches:

$$11\ 8\ 10\ 9\ 7\ . \qquad (3)$$

The retrograde inversion of (1) (Fig. 1G) is obtained by subtracting the pitches from a constant and reversing their order—that is, the reversal of (2):

$$1\ 4\ 2\ 3\ 5\ . \qquad (4)$$

Note that the inversion and retrograde transformations take one operation on the vector of pitches, while retrograde inversion takes two operations.

The description in terms of intervals relies on a characterization of a melody as a vector of signed interval sizes in semitones. (See Dowling & Fujitani, 1971, for more details.) For example, the intervals of Fig. 1A and Expression 1 are:

$$+2\ +1\ -2\ +3\ . \qquad (5)$$

The inversion may be derived by multiplying Expression 5 by −1:

$$-2\ -1\ +2\ -3\ . \qquad (6)$$

The retrograde inversion may be obtained by reversing the order of the intervals in (5). Observe that the retrograde inversion of the intervals is obtained by taking the retrograde of the pitches. Musicians use the pitch terminology rather than the interval terminology, and I will do the same here to avoid confusion. Experssion 7 represents the retrograde inversion of (5):

$$+3\ -2\ +1\ +2\ . \qquad (7)$$

The retrograde may be obtained by reversing the order of (5) and then multiplying by −1:

$$-3\ +2\ -1\ -2\ . \qquad (8)$$

Note that the retrograde of a melody takes two operations on the vector of interval sizes, while the inversion and retrograde inversion take only one. In actual musical practice and in the present experiment, these transformations are almost always accompanied by transposition, which in the pitch characterization amounts to adding a constant to the pitches and in the interval characterization amounts to selecting a different starting point.

Some of the transformed stimuli used in the present study are transforms of the contour (the pattern of ups and downs) of the melody, without preserving the exact interval sizes of the original. In Fig. 1, Examples D, F, and H are such contour transformations. They are most conveniently described as preserving the vectors of signs of Expressions 6, 7, and 8, but not the magnitudes. Thus the contour inversion of (5) is:

$$-\ -\ +\ -\ . \qquad (9)$$

Contour transformations are included along with exact transforms in this study because in actual musical practice these transforms, at least in the two kinds of inversion, tend to be contour-preserving, since exact interval size preservation would destroy the tonal scale relationships of the original (Apel, 1944). Second, Dowling (1971) showed that in short-term recognition memory for inversions, the contour seems to be all that is remembered. That result is checked here with retrogrades and retrograde inversions as well as inversions.

A question raised by the present study is: Which of the above characterizations of the process of transforming a melody is more like what listeners do when they recognize transformed melodies? Plausible arguments can be marshaled on both sides of the question. In favor of operations on vectors of interval sizes, there is (A) the fact that melodies are not changed by transposition, at least in long-term memory. The fact that melodies become less recognizable with transposition in short-term memory has been explained in two ways: first, Dowling and Fujitani (1971) argue that nontransposed melodies are easily recognized (in distinction to their particular confusion set) simply on the basis of their (unordered) set of pitches and not the pitch relationships which make up the "melody." Second, Francès (1958) and Dowling[1] argue that, at least for tonal melodies, a scale and a tonality provide a framework of pitch functions for remembering the melodic pitch relationships which carries over from standard to comparison stimuli. (B) The importance of contour for short-term memory noted above makes the interval size characterization, which codes contour directly, more attractive.

In favor of operations on vectors of pitches is (A) musicians' tendencies to think in terms of pitches (or images of pitches represented in notation) when describing the transformations—their use of descriptions like "upside down" for inversion and "upside down backwards" for retrograde inversion—which recall the operations on the pitch vectors outlined above. Sometimes this tendency to operate on the notes themselves has led composers to write so that turning the page upside down would lead to a viable retrograde inversion, or reading each line of music backwards a viable retrograde. (B) Inversion and retrograde, the simpler transformations of pitches, are the more prevalent transforms in Western music, while retrograde inversion, a simpler transform than retrograde in the interval characterization, is comparatively rare.

The results of the present study bear directly on the question of which kind of process listeners use when they recognize these transforms. If they use transformations on intervals, then retrograde should be the hardest to recognize; if they use pitch transformations, then retrograde inversions should be hardest. It is beyond the scope of this study to distinguish between processes which operate on pitches as a vector of pitch numbers and processes which operate on images of pitches (as of notes on a page). I will not try to maintain that the latter is the case, but only that in view of listeners' and my own introspections it seems highly plausible. Ss' reports of mental manipulation of the shapes sound very similar to the mental rotations of visual shapes described by Shepard and Metzler (1971).

The present experiment utilizes a short-term recognition-memory paradigm in which the comparison stimulus is one of the three transforms defined above of the standard. Separate groups of listeners performed the task with different transformations. I expected that (1) performance would be better than chance, thus demonstrating the possibility of the transforms' being perceived in music. (2) Inversions should be more easily recognized than the other transforms, since these are most common in music and both of the theoretical formulations above place inversions among the more easily handled perceptually. Whether retrogrades or retrograde inversions are easier should help decide whether the pitch vector or the interval vector, respectively, represents the more nearly correct description of the listener's behavior. (3) The issue of whether the exact interval sizes or just the contours are preserved in these transformations is tested by presenting both exact and contour transforms to all groups of listeners. Half of the groups had instructions to respond positively only to exact transforms, the other half to all contour transforms including the exact ones. Any tendency of Ss to differentiate between these two types of transforms should show up either as a main effect of the comparison-stimulus types (exact vs contour) or as an interaction of instructions (exact vs contour) with the stimulus types.

(4) In pilot studies, I noticed that Ss found the task extremely difficult, especially in the retrograde-inversion condition. Therefore, I devised a more extensive set of instructions involving an analogous visual task. Three groups of Ss did the visual task before performing the rest of the experiment. The visual task consisted of a recognition test of transforms of visual forms similar to notational representations of the melodies in the auditory experiment. I expected that doing this visual task would lead to improved performance on the auditory task. The main function I saw in the visual task was as a set of improved instructions communicating more effectively to the listeners the definitions of the various transforms to be recognized. (For example, until they actually worked it out in the ineluctable visual modality, many Ss found it hard to believe that the retrograde inversion of the contour of a monotonically ascending melody was itself a monotonically ascending melody.)

METHOD

The experiment was organized into a 5 conditions by 3 transforms by 2 stimulus types factorial design. Orthogonal comparisons were planned to test differences among the conditions. The first four conditions constituted a 2 by 2 factorial design of instructions [exact recognition (E) and contour recognition (C)] vs rate [fast (F) and slow (S)]. Planned comparisons were designed to test the main effects of instructions and rate. The first four conditions were all caried out without the analogous visual task preceding the recognition task. The fifth condition (VEF) was a replication of the exact-recognition fast condition (EF) but included the visual task. The effect of the visual task was tested by comparing Condition VEF with the other four conditions in a planned orthogonal comparison and also informally checking against Condition EF. The three transforms were inversion (I), retrograde (R), and retrograde inversion (RI). The two stimulus types were exact-interval size preserving and contour preserving. This last variable proved ineffectual, and the results discussed below are presented as collapsed across it.

Fifteen groups of Ss served in the 15 cells of the 5 conditions by 3 transforms design. Each group performed the recognition-memory task with the two types of comparison stimuli: transforms preserving the exact interval sizes of the standard and transforms of the contour of the standard. Positive recognition (hit) rates for each of these comparison types were plotted against false-positive response rates to random, different comparison stimuli for purposes of calculating areas under the memory operating characteristic (MOC) as an estimate of percentage correct in a two-alternative forced-choice task (see Norman & Wickelgren, 1965, for details). That is, all groups were scored the same regardless of instructions. Each group had 15 trials of each stimulus type—exact transform, contour transform, and different—making 45 trials in the session.

Subjects

There were 355 UCLA undergraduates from introductory psychology courses assigned to the 15 groups. Groups varied in size due to scheduling problems, and the smallest group contained 9 Ss. For this reason, an unweighted-means analysis of variance model was used. Ss came from a population shown in previous studies (Dowling, 1971; Dowling & Fujitani, 1971) to have a median amount of musical experience of 1.0 year and a mean of 2.25 years. Groups were assigned randomly to conditions.

Stimuli

A Hewlett-Packard 2116B computer generated the stimuli, which consisted of sawtooth waves. Stimuli were recorded on tape and played to Ss over high-quality sound-reproduction equipment. On each trial, the computer generated a new five-note standard stimulus by starting on middle C (262 Hz) and selecting succeeding notes such that intervals between successive notes occurred with the following probabilities: $P(\pm 1$ semitone$) = .5$, $P(\pm 2$ semitones$) = P(\pm 3$ semitones$) = .25$. The comparison melody started on a different note selected at random from the 14 notes of a chromatic scale 1-7 semitones higher or lower than middle C. Stimuli were presented at rates of 5 tones/sec (Conditions EF, CF, and VEF) or 2 tones/sec (Conditions ES and CS). Comparison stimuli followed standard stimuli by 2 sec. Ss had 5 sec to respond. A warning tone (4,250 Hz) preceded the onset of each trial by 2 sec.

Exact-interval size-preserving comparison stimuli were exact inversions, retrogrades, and retrograde inversions of the standard stimuli, each beginning on a new note. Contour-preserving comparisons contained newly selected intervals, different from the corresponding intervals of the standard but with the same contour transformed in the appropriate way, beginning on a new note. Different comparisons were selected just as the standards, but with the constraint that the contour be different in at least one interval direction from the appropriate transform of the standard.

New selections of standards were made randomly by the computer for each of the 15 experimental groups. Each new standard was constructed independently of all previous standards.

Procedure

On each of the 45 trials of the experiment, S heard a standard melody followed by a comparison melody. S's task in Conditions EFI and ESI was to say whether the comparison melody was an exact

Table 1
Areas Under the MOC Averaged Across Types of Stimuli

Transforms	Condition					Mean
	EF	ES	CF	CS	VEF	
Inversion	.67	.70	.70	.80	.70	.70
Retrograde	.54	.79	.59	.67	.65	.64
Retrograde Inversion	.53	.53	.50	.68	.60	.55
Mean	.59	.64	.61	.73	.65	

inversion of the standard or not. Ss responded on a four-category confidence level scale with categories labeled "sure same," "same," "different," and "sure different." The response "same" in this case meant that the comparison stimulus was an exact inversion of the same melody as the standard. No feedback was given Ss. The tasks in Conditions EFR, ESR, EFRI, and ESRI were similar except that Ss in those conditions had to recognize exact retrogrades and retrograde inversions. These conditions were preceded by instructions explaining the nature of the transform to be recognized and the importance of distinguishing between exact transforms and merely contour-preserving transforms. Three samples of each of the three trial types—exact, contour, and different—accompanied the instructions.

Ss in Conditions CFI, CSI, CFR, CSR, CFRI, and CSRI did a similar task to Ss in the other conditions, except that instead of recognizing exact transforms they had only to recognize transforms of contour. Instructions to these Ss explained the nature of the transform to be recognized and the concept of the countour-preserving transform as well as of the exact transform, but with the added stipulation that for the purposes of this experiment the two were to be considered equivalent. Again, sample trials accompanied the instructions.

Ss in the three VEF conditions did the same task as those in the EF conditions, but with the whole session initiated with a visual recognition task analogous to the auditory task. The task consisted of 10 problems. For each problem, there appeared at the left-hand side of the page a visual form (standard stimulus) consisting of five dots equally spaced from left to right of the standard, and, separated from it by a vertical rule on the page, were four comparison stimuli constructed in a manner similar to that of the standard. One of these comparisons was the appropriate transform of the standard, and S's task was to select the right one and put its number in the right-hand margin of the page. The problems included all the various contour shapes made with five points and their inversions. All transforms were exact. Ss were given feedback as to their correctness after the 10 problems had been completed. Ss then proceeded to the main part of the experiment.

RESULTS

In the 5 by 3 by 2 analysis of variance, the main effects of conditions [$F(4,680) = 3.73$, $p < .01$] and transforms [$F(2,680) = 11.10$, $p < .001$] were significant. The descending order of difficulty was inversion, retrograde, and retrograde inversion. The main effect of stimulus types was not significant, and, for purposes of further description, the data are collapsed across stimulus types. The only significant effect involving stimulus types was a Stimulus Types by Transforms interaction [$F(2,680) = 4.48$, $p < .01$] due almost entirely to the easiness of the contour stimuli in the R conditions, especially the RS condition—overall a difference of 71% correct vs 57% correct. Whether this is to be attributed to the ease of recognizing retrograde contours at slow speeds or to the nature of the particular stimuli generated in this experiment is not settled. The planned comparison of fast vs slow rates was significant [$F(1,680) = 4.25$, $p < .05$], with the slow rate easier (60% vs 68% correct). The planned comparisons of instructions and visual task were nonsignificant. In order to check further on the effect of the visual task, I made a post hoc test on the difference between the VEF and EF conditions. The difference, although in the predicted direction (65% vs 59% correct), was not significant.

Table 1 shows the data collapsed across stimulus types so that, in effect, the "same" responses to both exact and contour stimuli are counted correct. It is evident that the ascending order of difficulty of transforms is I, R, RI; and that this is roughly true of all but the ES conditions. All of the slow presentation rate groups are superior to their corresponding fast groups, except ESRI, which is equal to EFRI. There is little difference between groups EFI and VEFI but greater difference due to the visual task in the VEFR and VEFRI groups.

Inversions were recognized with better than chance accuracy under all conditions. Retrogrades were recognized better than chance in all except Condition EF. Retrograde inversions were recognized better than chance only in the CS and VEF conditions.

DISCUSSION

These results clearly demonstrate that inversions, retrogrades, and retrograde inversions of brief melodies can be recognized with better than chance accuracy. The most troublesome of these, the retrograde inversion, was recognizable in one condition at the slow rate and was recognizable at the fast rate when the session was preceded by the visual task clarifying the notion of retrograde inversion. Retrogrades were recognized with better than chance accuracy in all but Condition EF. Although the present experiment presented brief melodies in isolation from any confusing background and in that way made them easier to recognize than they would have been in an actual musical context, there are several reasons why recognition in the experiment should be more difficult than in music. (1) Atonal melodies such as those used here are typically difficult to deal with in recognition experiments because of their departure from the well-learned scale functions the listener hears in the rest of his musical experience (Francès, 1958). (2) The intervals used in the present stimuli are much smaller than those encountered in normal melodies. The median interval size of one semitone is much smaller than the median of about three semitones found in most melodies (Dowling & Fujitani, 1971). (3) In actual music, there is a rhythmic dimension present that was avoided in this experiment. White (1960) found that familiar melodies could be recognized with better than chance accuracy on the basis of their rhythm alone. Presumably rhythmic differences in actual music would serve to differentiate among alternatives, not only for inversions in which rhythmic patterns remain unaltered, but also in the retrogrades in which characteristic clusters of longer or shorter notes would still be distinguishable. (4) In the construction of a piece of music, composers usually choose melodic material so that separate melodies will be clearly distinguishable from each other. In recognition experiments like the present one, however, materials are selected to provide maximal homogeneity. Thus, the set of alternative stimuli in music is divided into a few clearly defined sets with corresponding responses. This situation is made explicit in Francès's (1958) Experiment X, in which Ss listened to a Beethoven scherzo, tapping once each time they heard the first theme and tapping twice each time they heard the second theme. In contrast to Francès's situation with stable stimulus classes distinguished along several dimensions at once, stimuli in the present experiment were distinguished on only a few dimensions and the positive stimulus set changed on each trial.

Dowling's (1971) result that exact interval size information becomes lost in recognition of inversions was replicated and extended to retrogrades and retrograde inversions. The only exception to the failure of either contour stimuli or contour instructions to make a difference in Ss' behavior compared with exact stimuli and instructions was the comparative easiness of the retrograde stimuli in the slow conditions. This could be because the particular stimuli generated by the computer in those conditions were easier to recognize in their transformations. (In my opinion, linear ascending and descending forms were easily recognized, but such an opinion would require checking by a separate experiment.) It is a defect of the present design that the computer was left to generate each new standard independently rather than select it from a controlled population. I attribute the unexplained variance here to lack of control over the standard stimulus forms and suggest that future experiments exert more control.

Slow presentation conditions were easier than fast. Further work is needed to decide whether the advantage of slow presentation lies chiefly in Ss' being better able to store the stimulus when he hears it or in the added time he has to manipulate the material. Future experiments should explore the relative effects of changing presentation rates within stimulus patterns and of changing interstimulus intervals.

Inasmuch as retrograde inversions were the most difficult to recognize, the pitch-vector characterization of the process by which Ss handle the task seems the more plausible psychological model than does the interval-vector characterization. Ss often described their method of doing the transformations in terms of reversing an image of the notes, or of turning it upside down—descriptions compatible with the pitch-vector model. Thus, Ss' introspections agreed with those of most musicians. However, some reservations about this conclusion should be noted. The retrograde and inversion transforms were not equally well recognized, as the pitch-vector model would predict. As good a case for the easiness of inversions as opposed to the other two transformations could be made as for the difficulty of retrograde inversions. Shepard[2] has suggested that the inversions might be easier because among the three they allow for the element-by-element comparison of the comparison stimulus with memory of the standard in the same temporal order. Suppose the standard is coded as in Expression 9: "Down, Down, Up, Down." Then, if S hears a stimulus going "Up, Up, Down, Up," he need only compare the two, using the rule that for each "Up" there should be a corresponding "Down" in memory, and vice versa. This kind of procedure will not work for retrogrades and retrograde inversions. In those cases, the order of one of the stimuli must be reversed at some point and the comparison made with one reversed and one nonreversed stimulus. Reversal can only occur after storage of the whole stimulus has occurred. Thus, in this view, performance with the retrogrades and retrograde inversions would be considerably complicated over that for inversions. In this view, the increased difficulty of retrograde inversions over retrogrades would be explained simply by the added complication of having to perform the inversion transformation—an operation that causes a significant decrement in performance over straight melody recognition (Dowling, 1971). In such a view, the temporal dimension would take precedence over pitch in importance, in the sense that its disturbance is more disruptive of recognition behavior. When Dowling's (1971) data are used as a baseline, performance is seen to drop from 87% correct to 70% or 73% with pitch inversion. However, performance drops from 87% to 64% with temporal reversal. The tendency of some contemporary music theorists to view the pitch and duration domains as analogues to the two dimensions of the painter's canvas seems to run into difficulties with the nature of musical perception.

REFERENCES

APEL, W. *The Harvard dictionary of music.* Cambridge, Mass: Harvard University Press, 1944.

DAVID, H. T., & MENDEL, D. *A Bach reader.* New York: Norton, 1966.

DOWLING, W. J. Recognition of inversions of melodies and melodic contours. Perception & Psychophysics, 1971, 9, 348-349.

DOWLING, W. J., & FUJITANI, D. S. Contour, interval, and pitch recognition in memory for melodies. Journal of the Acoustical Society of America, 1971, 49, 524-531.

FRANCÈS, R. *La perception de la musique.* Paris: J. Vrin, 1958.

NORMAL, D. A., & WICKELGREN, W. A. Short term recognition memory for digits and pairs of digits. Journal of Experimental Psychology, 1965, 70, 479-489.

RESTLE, F. Theory of serial pattern learning: Structural trees. Psychological Review, 1970, 77, 481-495.

SHEPARD, R., & METZLER, J. Mental rotation of three-dimensional objects. Science, 1971, 171, 701-703.

TOVEY, D. F. *The forms of music.* Cleveland: Meridian Books, 1956.

WHITE, B. Recognition of distorted melodies. American Journal of Psychology, 1960, 73, 100-107.

NOTES

1. Dowling, W. J. Musical scales as cognitive structures. In preparation.

2. R. Shepard, personal communication, 1972.

(Accepted for publication April 3, 1972; revision received July 2, 1972.)