

The interaction of modality condition and presentation rate in short-term contour recognition

WILLIAM R. BALCH and DENNIS L. MUSCATELLI
Pennsylvania State University, Altoona, Pennsylvania

Using a short-term recognition technique, musically experienced and musically inexperienced subjects indicated whether or not the contour of a probe, consisting of a sequence of six tones (auditory) or of a marker changing positions on each of six vertical lines (visual), was the same as that of an auditory or visual target presented just before. Thus there were four modality conditions: auditory target-auditory probe (A-A), auditory target-visual probe (A-V), visual target-auditory probe (V-A), and visual target-visual probe (V-V). All subjects were tested at five different presentation rates: $1/10$, $1/4$, $1/2$, 1, and 2 sec per note/position. For the A-V, V-A, and V-V conditions, recognition accuracy increased as presentation rate became slower. However, for purely auditory recognition (A-A), accuracy increased more slowly (for musically experienced subjects) or not at all (for inexperienced ones). This interaction of modality condition and rate is discussed in terms of its implications for the abstraction of contour from melody.

Melodic contour refers to a particular sequence of up/down pitch changes in a series of notes, regardless of the exact intervals (i.e., frequency ratios) between adjacent notes. A number of studies have revealed that such contour information figures importantly in the immediate recognition of short and novel melodies (e.g., Bartlett & Dowling, 1980; Dowling, 1978; Dowling & Bartlett, 1981; Dowling & Fujitani, 1971; Idson & Massaro, 1978; and see also Dowling, 1982, and Edworthy, 1982, for a discussion of boundary conditions related to the effectiveness of contour in melodic memory). Typically, these studies have shown that a probe melody with different intervals and pitches from those of a target item presented just before will generally be mistaken for the original if the contour has been preserved. However, if the probe has an altered contour, it can be better discriminated as being different from the target. This effect does not require musical experience on the part of the subject. Indeed, Trehub, Bull, and Thorpe (1984), in their Experiment 2, have shown essentially the same effect in infants aged 8-11 months, using the technique of visually reinforced head-turning as a measure of discrimination.

Several studies have demonstrated the apparent similarity between the perceptual and memory characteristics of melodic contour and those of visual contour. One such similarity concerns the salience of particular contour fea-

tures. For instance, the angles or corners of a visual figure are considered to have high information value (Attneave, 1954), and they are fixated by the subjects longer than other parts (Baker & Loeb, 1973). Accordingly, Dyson and Watkins (1984) tested the salience of melodic contour reversal points (an up directional change followed by a down, or vice versa), which are analogous to visual corners. The listeners in Dyson and Watkins's study discriminated between same and different melody pairs more accurately if the pitch changes occurred at contour reversals rather than at nonreversal points (at which the melody did not change direction but continued up or down).

A second example of the comparability of visual and auditory contour is found in the phenomenon of stream segregation. Primary auditory stream segregation (Bregman, 1978; Bregman & Campbell, 1971; Bregman & Rudnicki, 1975) occurs when a series of alternating high- and low-frequency tones is played. If the presentation rate is slow, these notes sound like a single continuous contour, or stream. However, at faster rates, two contours can be heard, one consisting of only the high tones and the other consisting of only the low. Likewise, visual stream segregation (Achim & Bregman, 1973) is found when a series of lights, some high and some low in the visual field, are alternated at a sufficiently rapid rate. This analogy between visual and auditory stream segregation is bolstered by O'Leary and Rhodes's (1984) finding that when both auditory and visual sequences are presented at the same time, the perception of two streams in one modality influences the presentation rate required to produce stream segregation in the other modality. This cross-modal effect indicates perceptual similarities between the auditory and visual displays used in that study.

Another type of cross-modal influence, which relates to contour interference with melodic short-term memory

This work was supported by a Pennsylvania State University Faculty Development Grant to the first author. The authors wish to acknowledge the helpful comments of W. J. Dowling, Robert Crowder, and Andrea Halpern, and also the assistance of Rhonda Stern and Norma Ritz in graphic work and manuscript preparation, respectively. The mailing address of the authors is: Department of Psychology, Pennsylvania State University, Altoona, PA 16603. Reprint requests should be sent to William R. Balch at that address.

code, was found by Balch (1984). Subjects were required to recall, immediately after presentation, the sequence of up/down pitch changes in the contour of seven-note melodies. The recall tasks were designed to interfere with the memory code of the melody with respect to either contour or modality.

Contour interference proved to be more important than modality-specific interference. In one of Balch's (1984) experiments, for example, three factors—contour interference, modality of recall task, and presentation mode—were manipulated separately. Contour interference was found to increase the recall time of both visual and auditory contours, regardless of the modality of the recall task. This finding suggests that visual and auditory representation of contour have a common abstract component.

Balch (1984) also found some effects of presentation modality. In an experiment that compared recall of visual and auditory contours, the former were consistently recalled faster and more accurately than the latter. Additionally, subjects were only slightly slower when they had to recall the inversion of a visual contour (i.e., the mirror-image or upside-down version) than when they were free to recall the contour without having to invert it. However, subjects hearing auditory contours were much slower on inverted than on noninverted recall. These findings appear to reflect differences in the efficiency with which contour is abstracted from a visual as opposed to an auditory stimulus.

In the present study, the process of contour abstraction was explored further, using the method of short-term recognition. With this technique, the target stimulus could be either auditory or visual. The auditory target consisted of a six-note melody; in the visual display, a horizontal marker successively changed vertical positions on a series of six vertical lines spaced across a CRT screen. After presentation of the target stimulus and an interval of 2 sec, another stimulus, called the probe, was presented. The probe could also be either auditory or visual. Sometimes the probe had exactly the same contour as the target, and sometimes it was altered at one point in the contour (up instead of down, or vice versa). The subject's task was to indicate whether the probe had the same contour as the target or a different one. Note that four modality conditions were generated: auditory target-auditory probe (A-A), auditory target-visual probe (A-V), visual target-auditory probe (V-A), or visual target-visual probe (V-V). In addition, subject groups of both high and low musical experience were employed.

The hypothesis tested here will be called the *contour abstraction efficiency hypothesis*. It maintains that up/down contour abstraction is more efficient for visual presentation than it is for auditory presentation. To test this hypothesis, it was assumed that the process of abstraction requires time. Thus, contour recognition accuracy should be difficult when presentation rate is extremely fast, regardless of modality condition. Contour

abstraction efficiency can be assessed, for each condition, by considering the improvement in recognition accuracy as presentation rate becomes slower. To this end, five presentation rates were employed, ranging from $1/10$ sec to 2 sec per note or visual position.

According to the above hypothesis, unimodal visual recognition accuracy (V-V) should increase markedly as presentation rate becomes slower. In other words, V-V subjects should be best able to take advantage of the extra time available to abstract contour. However, accuracy of unimodal auditory recognition (A-A) should improve more slowly—or not at all—as presentation rate is slowed, reflecting less efficiency of contour abstraction. Thus there should be an interaction between presentation rate and modality condition.

Since response probability is the dependent measure, it is important to consider the nature of the relationship between this measure of recognition accuracy and the underlying construct of contour abstraction. Otherwise, problems can arise in interpreting interactions between independent variables. It is assumed here simply that the probability of a correct recognition response is an increasing monotonic function of the quality of contour abstraction. Under this assumption, only interactions of ordinal nonindependence are interpretable. Such interactions, also called crossover interactions, occur when the curves of two independent variables plotted against each other touch (Loftus, 1978).

The question of how subjects might perform in the cross-modal recognition conditions (A-V and V-A) is interesting. Suppose subjects do abstract contour less efficiently under auditory than under visual presentation, as the present hypothesis states. This disadvantage could arise either (1) because they have a limited ability to abstract melodic contour or (2) because they do not abstract contour as spontaneously from melodies as they do from visual patterns.

If explanation 1 is correct, then both A-V and V-A recognition should follow the pattern of the A-A condition. Recognition accuracy should improve little or not at all as presentation rate decreases, since poor ability to abstract auditory contour would serve as a limiting factor in the accuracy of either cross-modal condition. However, if explanation 2 is correct, recognition accuracies in both cross-modal conditions would be likely to increase substantially with slower presentation rates. Subjects might show poor performance in purely auditory A-A recognition regardless of presentation rate. However, the cross-modal conditions should more effectively force attention to contour. Therefore, cross-modal recognition should be relatively accurate at rates slow enough to allow contour abstraction. Since the possibility for comparing the sensory characteristics of target and probe items is eliminated in the cross-modal conditions, abstraction of up/down contour now becomes the primary basis on which a comparison of the auditory and visual stimuli would be made.

METHOD

Subjects

Ninety-six male and female students enrolled in the introductory psychology course at Pennsylvania State University, Altoona, participated in this experiment. All received a small amount of academic credit for their participation.

Half of the subjects had been designated, on the basis of a questionnaire, as low in musical experience. The subjects in this category had had no musical experience or had played, or taken lessons on, a musical instrument for less than 1 year. In addition, each student reported that he/she was unable to read music.

The other half were classified as high in musical experience. They had taken at least 4 years of lessons on a pitched musical instrument, played currently, and had casually or formally performed on their instruments for others within the previous 2 years. In addition, each student reported that he/she was able to read music.

Design

A three-way mixed factorial design was used. Subjects of two musical experience levels (high and low) were assigned to one of four modality conditions: auditory-auditory (A-A), auditory-visual (A-V), visual-auditory (V-A), or visual-visual (V-V). A total of eight groups were thus generated, consisting of 12 subjects each. Every group was tested for contour recognition at each of five presentation rates: $\frac{1}{10}$, $\frac{1}{4}$, $\frac{1}{2}$, 1, and 2 sec per note (auditory contour) or vertical position (visual contour).

Materials and Apparatus

Auditory contours. Melodic contours consisted of six sawtooth-wave tones generated by a Commodore 64 microcomputer. The sound interface device in the microcomputer was a chip programmable from 30 Hz to 12 kHz with a resolution of 1 Hz. Contours were heard by subjects at approximately 60 dB SPL on the loudspeaker of a Commodore 1702 monitor. For each tone, attack time (interval from onset to maximum intensity) was 2 msec; there was no decay, since the tone was sustained at maximum amplitude; and release time (time from maximum amplitude to offset) was 6 msec.

Twelve auditory target contours were employed in the A-A and A-V conditions. The pitches in each are specified in the column of Table 1 labeled "Target." Each contour started with the pitch C5 (524 Hz) and then underwent five random up or down pitch changes. This random sequence of directional changes is shown, for each target, in the "Base Contour" column. Intervals of pitch change were always one diatonic step, and were equally likely to be up or down. Target contours could be likened to a six-note melody played on the white keys of a piano, starting with high C and then proceeding up or down by one white key at a time.

Auditory probes, used in the A-A and V-A conditions, began two diatonic steps higher (at E5 = 659 Hz) or lower (at A4 = 440 Hz) than the starting point for targets. In Table 1, the "Base Contour" column indicates whether the shift from target to probe was up (+) or down (-). It also shows where in the contour the direction of pitch change was altered on the half of the trials when the probe was different. For instance, for contour 1 (DUUUD), the second direction (underlined in the table) is altered, producing a different contour (DDUUD). The resulting pitches are specified in the "Different Probe" column. Only the three middle positions of contour were ever changed for different probes, since changes at the beginning or end of a contour might be too noticeable. The probe contained an altered contour on half of the trials.

On the other half of the trials, the contour of the probe was unchanged from that of the target (see "Same Probe" column of Table 1). Although these probes preserved contour, the pitches (i.e., frequencies) were shifted, and in no case were all the exact intervals (i.e., frequency ratios rather than diatonic intervals) identical for target and probe. Thus, in A-A recognition, the subjects could

not detect "same" probes on the basis of exact melodic transposition. This consideration is important because otherwise A-A subjects could utilize an extra cue not available to the other three modality recognition conditions.

On the other hand, it could be argued that not transposing melodies exactly from target to probe might actually confuse A-A subjects. For instance, subjects might judge a "same" probe as being different if they base their judgments on exact transposition rather than contour. In short-term recognition, however, novel melodies are not distinguishable with above-chance accuracy from "tonal answers," which are probes that preserve both contour and the number of diatonic steps in each interval (Dowling, 1978). Therefore, it should make no difference whether the probes are exact transpositions, or, as in the present case, tonal answers.

Visual contours. The scheme illustrated for auditory contours in Table 1 applies to visual ones as well. Figure 1A illustrates a visual contour that is the counterpart of auditory target 1 (DUUUD). Visual target contours were used in the V-V and V-A modality conditions. The vertical position of a single horizontal bar was changed on each of a series of six vertical lines, starting halfway down the leftmost line and then proceeding up and down and to the right. Although in the figure all the vertical positions are shown, in the actual target or probe displays, only one horizontal bar was visible as it changed positions on each successive vertical line. The six vertical lines were visible throughout the entire presentation of the contour. All visual contours were generated by a Commodore C-64 microcomputer and viewed on a Commodore 1702 monitor. The contour display, unlike that shown in Figure 1, was white on a dark background. The horizontal bar subtended a width of 2.58° and a thickness of $.5^\circ$. The vertical lines were 15.83° high, $.25^\circ$ thick, and spaced 3.83° apart. Changes in vertical position made by the marker were 1.42° up and down.

Visual probes, used in the V-V and A-V conditions, are illustrated in Figure 1B. These probes were positioned 2.84° higher or lower on the CRT screen than the visual targets—twice the distance between vertical marker positions within a contour. This procedure was analogous to that of making the auditory target-to-probe shift (two diatonic steps) twice the interval between notes within a contour (one diatonic step). Figure 1B shows a "different" probe. Since it was based on target contour 1 (see Table 1), the shift was upwards and the second directional change was altered. Thus, the contour was now DDUUD instead of DUUUD.

All contours, whether visual or auditory, were based on single steps between positions or notes, and were designed to be perceived as a single stream regardless of presentation rate. No stream segregation was perceived by pilot subjects or by the experimenters, nor was it reported by any subjects during the postexperimental interviews.

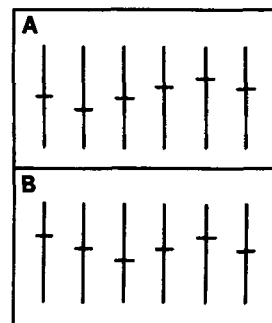


Figure 1. Sample visual contours used in the experimental trials. Box A illustrates a target, and Box B, a probe contour. Although all the positions of the marker are shown in this figure, in the actual contour displays, a single marker changed positions from left to right at one of five rates.

Table 1
Auditory Contours Used in Experimental Recognition Trials

Base Contour	Target	Same Probe					Different Probe						
1. <u>D</u> UUUD(+)	8 7 8 9 10 9	10	9	10	11	12	11	10	9	8	9	10	9
2. U <u>D</u> UDD(+)	8 9 8 9 8 7	10	11	10	11	10	9	10	11	12	13	12	11
3. UD <u>U</u> DU(+)	8 9 8 9 8 9	10	11	10	11	10	11	10	11	10	9	8	9
4. UUD <u>U</u> D(+)	8 9 10 9 10 9	10	11	12	11	12	11	10	11	12	13	14	13
5. UDD <u>U</u> D(+)	8 9 8 7 8 7	10	11	10	9	10	9	10	11	10	9	8	7
6. DUU <u>D</u> D(+)	8 7 8 9 8 7	10	9	10	11	10	9	10	9	10	11	12	11
7. <u>D</u> UDDU(-)	8 7 8 7 6 7	6	5	6	5	4	5	6	5	4	3	2	3
8. U <u>D</u> DDU(-)	8 9 8 7 6 5	6	7	6	5	4	3	6	7	8	7	6	5
9. DD <u>U</u> DU(-)	8 7 6 7 6 7	6	5	4	5	4	5	6	5	4	3	2	3
10. DD <u>U</u> UU(-)	8 7 6 7 8 9	6	5	4	5	6	7	6	5	4	3	4	5
11. UDD <u>U</u> U(-)	8 9 8 7 8 9	6	7	6	5	6	7	6	7	6	5	4	5
12. DDD <u>U</u> U(-)	8 7 6 5 4 5	6	5	4	3	2	3	6	5	4	3	4	5

Note—A (+) indicates an upward shift from target to probe, and a (-) indicates a downward one. The underlined direction is the one that was changed for "different" probes based on that target. The letter names of notes and frequencies (in hertz) that correspond to the numerical designations in the table are: 1 = C₄ = 262; 2 = D₄ = 294; 3 = E₄ = 330; 4 = F₄ = 349; 5 = G₄ = 392; 6 = A₄ = 440; 7 = B₄ = 494; 8 = C₅ = 524; 9 = D₅ = 587; 10 = E₅ = 659; 11 = F₅ = 698; 12 = G₅ = 784; 13 = A₅ = 880; 14 = B₅ = 988.

Procedure

There were a total of 120 experimental trials, based on a combination of 12 targets, with two types of probe (same or different) presented at one of the five rates (specified in the Design subsection). Each trial consisted of a target and then a probe contour, presented at the same rate within the trial. Rate was defined by the duration of each tone or visual position in a contour. Within each contour, tones/positions were presented in immediate succession. There was, however, a 2-sec interval between target and probe contours. This length of interstimulus interval was selected in order to allow subjects a clear discrimination between the offset of the target and onset of the probe contour. The 2-sec interval is common in short-term melody-recognition studies (e.g., Dowling, 1978; Dowling & Fujitani, 1971).

Each participant was assigned randomly to one of the four modality conditions. The sessions were administered individually, and consisted of a warm-up and an experimental phase.

Warm-up phase. The subjects were given 10 trials to familiarize them with the requirements of contour recognition. Five contours, not used in the experimental trials, were selected as targets. They appeared in two trials each—once with a same probe and once with a different one. All five rates were illustrated in the warm-up. Half the trials employed upward probe shifts, and half employed downward shifts.

At the start of the warm-up trials, the subjects were told that they would hear (or see) a pattern of six notes (or visual positions), followed after a brief pause by another pattern. They were asked to determine whether the up/down contour of the second pattern matched that of the first, and to respond by pressing either the button marked "same" or the button marked "different." The subjects were told that their responses had to be made after the onset of the last note or visual position of the probe, and that otherwise the responses would not be registered by the microcomputer and would have to be made again. Response buttons were mounted about 5.5 cm apart in a small box held by the subject and connected by cable to the microcomputer. For half the subjects, the "same" button was on the right; for the other half, it was on the left.

The subjects were told that their same/different judgments should be based only on up/down contour and not on the starting point of each pattern (i.e., the first note or visual position). Also, the subjects were told to consider only the direction of the change in pitch (or vertical height) and not the amount of up/down change. This instruction was relevant mainly to the A-A condition. Although

single diatonic steps were always used as intervals between notes, the step could be either a semitone or a whole tone.

The fact that auditory contours use two different pitch intervals and visual contours use only one vertical interval could suggest that the auditory contours were not comparable to the visual ones. However, since diatonic scales are prevalent in Western cultural practice, it seemed more desirable to use single diatonic steps than a single pitch interval (i.e., frequency ratio), which is extremely rare in musical practice. Since whole-tone scales, for instance, are avoided by most composers because they sound unnatural, results obtained here with such scales would not be generalizable to contour recognition in a normal musical context.

Each trial was followed by a feedback message on the monitor that told subjects whether or not their response was correct. Incorrect trials were repeated, and the correct answer was explained if necessary. By the end of the warm-up procedure, all subjects reported a clear understanding of the task requirements.

Experimental phase. The subjects were then told that the experimental trials would begin, and that these trials were exactly the same as the warm-ups, except that no feedback about the correctness of the responses would be given. Furthermore, they were informed that probes were equally likely to be same or different.

The intertrial interval was 2 sec, and a break of approximately 2 min was given half-way through the trials. After the last trial, a brief postexperimental interview was conducted in which subjects were questioned about the strategies they had used in the recognition task.

RESULTS

A percent-correct-responses score was calculated at each of the five presentation rates for each of the 96 subjects. Every score was based on 24 experimental trials (12 same probes and 12 different), and exactly the same contours were involved for each of the rates. Then a 2×4×5 mixed analysis of variance (musical experience × modality condition × presentation rate) was applied to these accuracy scores.

It should be noted that the area under the memory operating characteristic (MOC) curve (e.g., Swets, 1973) would be the best choice for the dependent measure on

the basis of purely psychometric considerations. However, this use of signal detection theory would require a more demanding mode of response than might be desirable for the present study. Typically, MOC-based measures are obtained by having subjects respond on a 4-point confidence scale, such as: sure different, different, same, sure same (e.g., Dowling, 1978). The simpler, binary same/different response was selected here due to the potential difficulty of the stimulus materials used. Both target and probe were relatively long (six notes or visual positions). Moreover, both very fast and very slow presentation rates were included in the design; the former require especially careful attention, and the latter demand retention of the stimulus materials over quite a long period of time. Finally, the cross-modal comparisons involved in the A-V and V-A conditions represent a task requirement not typically encountered in short-term recognition tasks. Therefore, it was decided that the simplest possible response was required: the same/different judgment.

The experimental effects on recognition performance are illustrated in Figure 2. By comparing the leftmost graph in this figure (low subjects) with the center one (high subjects), it is apparent that musical experience facilitates contour recognition performance. The mean accuracy for low subjects was 73%, whereas for high subjects it was 82%. The main effect of musical experience was significant [$F(1,88) = 28.90, p < .001$], but there was no significant interaction of experience with modality condition [$F(3,88) = .99, p > .10$] or of experience with presentation rate [$F(4,352) = 1.39, p > .10$].

The result of main interest in this study is the significant interaction obtained between modality condition and presentation rate [$F(12,352) = 6.78, p < .001$]. [Each of these factors was also significant as a main effect: for modality, $F(3,88) = 10.94, p < .001$; and for rate, $F(4,352) = 96.32, p < .001$.] The interaction appears mainly to reflect the differences between the effect of presentation rate on A-A recognition and the effects of rate on the other modality conditions. Accordingly, separate analyses of the rate effects were performed on each condition. These analyses were additionally broken down for high versus low musical experience levels.

It should be noted that there was no three-way interaction of modality condition \times rate \times experience [$F(12,352) = .67, p > .10$]. However, it is reasonable to expect that musically experienced subjects might abstract melodic contour more efficiently than might inexperienced ones. If so, A-A contour recognition accuracy might increase—as presentation rate became slower—for musically experienced subjects, even if it did not for subjects of low experience. Since Figure 2 appears to conform to this trend, musical experience was considered in analyzing the modality \times rate interaction.

For musically inexperienced subjects, there was no increase in A-A recognition accuracy as rate became slower [$F(4,44) = .26, p > .10$]. However, for all other modality conditions, accuracy increased significantly with rate [for A-V, $F(4,44) = 18.39$; for V-A, $F(4,44) = 13.16$; and for V-V, $F(4,44) = 23.36$; all $ps < .001$]. The modality condition \times rate interaction calculated just

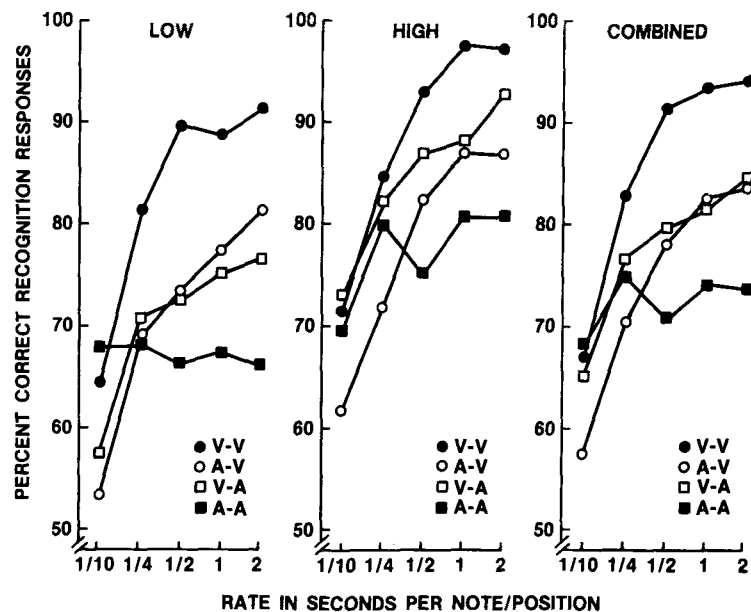


Figure 2. Mean accuracy scores as a function of presentation rate and modality condition. In the graphs, the modality of the target is indicated by the first A (for auditory) or V (for visual); the A or V following the dash indicates the modality of the probe. Results for subjects of low musical experience are shown in the leftmost graph; those for subjects of high experience are shown in the center graph. The results for all subjects combined are given in the rightmost graph.

for musically inexperienced subjects was reliable [$F(12,176) = 3.79, p < .001$].

Subjects high in musical experience also showed a clear modality \times rate interaction [$F(12,176) = 3.17, p < .001$]. In contrast to the result with inexperienced subjects, however, A-A recognition accuracy did increase significantly with rate [$F(4,44) = 4.78, p < .05$]. In all other modality conditions, accuracy increased significantly with rate [for A-V, $F(4,44) = 23.70$; for V-A, $F(4,44) = 13.16$; and for V-V, $F(4,44) = 22.47$; all $ps < .001$].

Another way of looking at the modality \times rate interaction is to compare A-A accuracy directly with accuracies for the other modality conditions when presentation rate is equal for each comparison. Two rates were chosen. One was the fastest ($1/10$ sec per note/visual position), at which contour abstraction should be most difficult. The other was the slowest (2 sec per note/position), at which contour should be most easily abstracted.

At the fastest rate, A-A recognition was significantly better than A-V performance and was not significantly different from V-A or V-V recognition. (See the rightmost graph of Figure 2, which shows the results for both musical experience levels combined.) The 95% confidence t intervals of the overall accuracy differences between the A-A conditions and A-V, V-A, and V-V conditions, respectively, were: $11.86\% \pm 6.68\%$, $4.48\% \pm 7.57\%$, and $1.28\% \pm 6.17\%$.

At the slowest rate, however, A-A accuracy was exceeded by every other modality condition. The corresponding confidence intervals for the differences between A-A accuracy and A-V, V-A, and V-V accuracies, respectively, were $8.17\% \pm 7.11\%$, $8.54\% \pm 7.62\%$, and $17.23\% \pm 6.66\%$. Thus, the relative difficulties of the modality conditions clearly depend on presentation rate.

DISCUSSION

The interaction between modality condition and presentation rate supports the contour abstraction efficiency hypothesis proposed here. At the fastest rate, unimodal visual (V-V) recognition was no more accurate than unimodal auditory (A-A) recognition. However, V-V accuracy improved more with slower rates than did A-A accuracy. Note that this interaction is one of ordinal nonindependence (Loftus, 1978). Therefore, it is interpretable, as discussed in the introductory section.

A modality \times rate interaction was also found in an earlier study by Garner and Gottwald (1968). However, they used quite different methods and stimuli. Their displays were eight-element sequences of binary visual or auditory events, presented repetitively. Moreover, they tested recall rather than recognition, and were not studying contour as such.

In the present study, the mode \times rate interaction is consistent with the conclusion that the abstraction of up/down contour is more efficient from visual patterns than from comparable melodic patterns. In other words, the V-V

subjects were more able than A-A subjects to make use of the extra time to better encode target and probe items into contour.

This conclusion is reinforced by another result, which relates to the difference between the performance of subjects with high versus low musical experience. For both experience levels, there was a reliable modality \times rate interaction. Highly experienced A-A subjects, however, showed a significant increase in accuracy as rate became slower, whereas their musically inexperienced counterparts did not. Thus, the efficiency of melodic contour abstraction appears to be related to musical experience.

The question of whether inefficient auditory contour abstraction stems from (1) a relative inability to abstract contour, per se, or from (2) not spontaneously abstracting the contour of a melody, is addressed by the performance of the cross-modal (A-V and V-A) subjects. For these subjects, recognition accuracy increased markedly with slower presentation rates, even for those with low musical experience. At the fastest rate, for which contour abstraction should be difficult under any modality condition, V-A accuracy was not significantly different from purely auditory recognition (A-A) accuracy. Moreover, A-V performance at the fastest rate was significantly worse than that of A-A subjects. Yet, at the slowest rate, for which contour abstraction should be easiest, performance in both cross-modal conditions significantly exceeded A-A performance.

These results support explanation 2, which assumes simply that listeners do not as spontaneously abstract contour from melodies as they do from visual patterns. Even though contour recognition was required in the A-A condition, this unimodal auditory recognition task would not force attention to contour as effectively as would the cross-modal tasks. When target and probe were presented in different modalities, contour became emphasized as the basis of comparison. Cross-modal listeners apparently found no problem in abstracting contour from an auditory target (A-V condition) or probe (V-A condition) at sufficiently slow rates. The presence of a visual pattern was apparently sufficient to facilitate the abstraction of contour from an auditory pattern. Contour abstraction was also efficient in the unimodal visual (V-V) condition. It can thus be concluded that subjects more spontaneously abstracted contour information from the visual than from the auditory stimuli.

The authors acknowledge that there may be boundary conditions on the present results. Displays in two different modalities are virtually impossible to make equivalent in every other potentially relevant respect. Therefore, the results may turn out to be limited to certain parameters of auditory or visual displays.

A major assumption in this study has been that auditory and visual presentations of contour are indeed comparable.¹ This assumption is justified by several studies discussed earlier (Achim & Bregman, 1973; Balch, 1984; Bregman & Campbell, 1971; Dyson & Watkins, 1984; O'Leary & Rhodes, 1984), and is supported here as well.

The A-V and V-A subjects had to compare contours in two different modalities, and yet performed significantly better than the A-A subjects at the slowest rate. This result is noteworthy because usually cross-modal comparisons in a short-term recognition task are more difficult than unimodal comparisons (e.g., Chase & Calfee, 1969). In the present study, the relatively accurate cross-modal contour recognition obtained at slower rates suggests that the subjects readily understood the comparison of the auditory and visual patterns on the basis of contour. Furthermore, the subjects' reports were consistent with this conclusion. During the postexperimental interviews, even the musically inexperienced subjects reported a clear grasp of the analogy between visual and auditory contour.

During these interviews, most subjects—regardless of modality condition or experience level—reported trying to use the following basic strategy. During the presentation of the target, they would attempt to abstract the up/down sequence direction by direction. Then, as the probe was presented, they would in some way rehearse or play back the contour of the target in synchrony with that of the probe, responding “different” if they detected a discrepancy. The subjects generally reported that they attempted this strategy at all rates, but that it became progressively easier to use as presentation rate decreased.

In reporting their particular methods of encoding contour, the subjects showed considerable individual differences. A few reported trying to say words like “up” and “down,” or “high” and “low,” to themselves. However, over 85% reported no verbal or covert labeling. One common strategy was to move a finger up and down. Other subjects said they “just thought” about the contour patterns. Several described their strategies in modality-specific terms, such as mental humming of melodies or viewing of visual contours. Most, however, did not.

Indeed, hypotheses involving modality-specific encoding do not account adequately for the present data. This notion was originally applied by Brooks (1968) in the context of interference with verbal versus visual memory codes. It could be applied here by assuming that subjects' playback of the target during probe presentation was strictly modality-specific. Therefore, in the unimodal auditory (A-A) condition, the probe might interfere with the memory of the target. Such interference explains the relatively poor A-A performance across rate. However, this account falls short for V-V recognition, where modality-specific interference should also occur. Note that V-V performance was higher than cross-modal performance (where no interference should occur) at all presentation rates (see the Combined graph in Figure 2). A viewpoint stressing modality-related interference should predict the opposite.

A dual-encoding account, attributing modality-specific encoding to auditory patterns but not to visual ones, is equally problematic. According to this view, A-A but not V-V recognition should produce interference at all presentation rates. At the fastest rate, however, A-A recognition was significantly better than A-V recognition. A dual-encoding viewpoint would predict the reverse result, since

no interference should be involved for the cross-modal A-V condition.

In conclusion, the modality \times rate interaction obtained here suggests a difference in the efficiency with which contour is abstracted from visual, as opposed to auditory, patterns. This difference appears to relate mainly to performance rather than to ability. That is, auditory contour can be abstracted more readily in cross-modal than in unimodal recognition. In the former case, subjects' attention is more effectively directed toward contour as abstract information. In other words, the presence of a visual pattern appears to facilitate the abstraction of contour from an auditory one. An interesting implication of this view is that cross-modal recognition might serve as a basis for training subjects to improve their subsequent memory for melodic contour. The present authors are currently working on the development and testing of such training methods.

REFERENCES

- ACHIM, A., & BREGMAN, A. S. (1973). Visual stream segregation. *Perception & Psychophysics*, *13*, 451-454.
- ATTNEAVE, F. (1954). Some informational aspects of visual perception. *Psychological Review*, *61*, 183-193.
- BAKER, M. A., & LOEB, M. (1973). Implications of measurements of eye fixations for a psychophysics of form perception. *Perception & Psychophysics*, *13*, 185-192.
- BALCH, W. R. (1984). The effects of auditory and visual interference on the immediate recall of melody. *Memory & Cognition*, *12*, 581-589.
- BARTLETT, J. D., & DOWLING, W. J. (1980). The recognition of transposed melodies: A key distance effect in developmental perspective. *Journal of Experimental Psychology: Human Perception & Performance*, *6*, 501-515.
- BREGMAN, A. S. (1978). The formation of auditory streams. In J. Requin (Ed.), *Attention and performance VII* (pp. 63-75). Hillsdale, NJ: Erlbaum.
- BREGMAN, A. S., & CAMPBELL, J. (1971). Primary auditory stream segregation and perception of order in rapid sequences of tones. *Journal of Experimental Psychology*, *89*, 244-249.
- BREGMAN, A. S., & RUDNICKY, A. I. (1975). Auditory segregation: Stream or streams? *Journal of Experimental Psychology: Human Perception & Performance*, *1*, 263-267.
- BROOKS, L. R. (1968). Spatial and verbal components of the act of recall. *Canadian Journal of Psychology*, *22*, 349-368.
- CHASE, W. G., & CALFEE, A. C. (1969). Modality and similarity effects in short-term recognition memory. *Journal of Experimental Psychology*, *81*, 510-514.
- DOWLING, W. J. (1978). Scale and contour: Two components of a theory of memory for melodies. *Psychological Review*, *85*, 341-354.
- DOWLING, W. J. (1982). Contour in context: Comments on Edworthy. *Psychomusicology*, *2*, 47-48.
- 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.
- DYSON, M. C., & WATKINS, A. J. (1984). A figural approach to the role of melodic contour in melody recognition. *Perception & Psychophysics*, *35*, 477-488.
- EDWORTHY, J. (1982). Pitch and contour in music processing. *Psychomusicology*, *1*, 44-46.
- GARNER, W. R., & GOTTFWALD, R. L. (1968). The perception and learning of temporal patterns. *Quarterly Journal of Experimental Psychology*, *20*, 97-109.
- IDSON, W. L., & MASSARO, D. W. (1978). A bidimensional model of

- pitch in the recognition of melodies. *Perception & Psychophysics*, **24**, 551-565.
- LOFTUS, G. R. (1978). On interpretation of interactions. *Memory & Cognition*, **6**, 312-319.
- O'LEARY, A., & RHODES, G. (1984). Cross-modal effects on visual and auditory object perception. *Perception & Psychophysics*, **35**, 565-569.
- SWETS, J. A. (1973). The relative operating characteristic in psychology. *Science*, **182**, 990-1000.
- TREHUB, S. E., BULL, D., & THORPE, L. A. (1984). Infants' perception of melodies: The role of melodic contour. *Child Development*, **55**, 821-830.

NOTE

1. It could be argued that a masking effect may have occurred with the auditory but not the visual contours. This point bears on the issue

of comparability between the two types of contour, and should be addressed. If auditory masking did occur, it should be most pronounced with the faster rates, where the tones are closer to the duration threshold for accurate recognition of pitch. However, masking should be minimal for the slower rates, where the tone durations are far above the pitch-recognition threshold. The masking argument should thus predict that recognition accuracy should increase for the A-A condition. The flat A-A function found for musically inexperienced subjects suggests that masking did not appreciably influence the results of this study.

(Manuscript received March 7, 1986;
revision accepted for publication August 20, 1986.)