Recognition time for excerpts from musical compositions

KATHY D. STEWART and W. A. WILBANKS
University of Georgia, Athens, Georgia 30602

Does the time required for listeners to recognize excerpts from musical compositions depend upon whether the excerpt is from the beginning, middle, or end of the composition? The musical compositions were four tape-recorded Debussy études for piano with distinct (nonrepetitious) beginning, middle, and end sections. The compositions were learned one at a time. The recognition tests determined response times for recognizing excerpts from the beginning, middle, and end sections. The results indicate that whether a musical phrase is at the beginning, middle, or end of a composition does not in itself determine recognition time. Although we have to hear the beginning of a composition before we hear the end, this temporal order in perception does not determine the cognitive structure of the composition as remembered.

Music is a temporal phenomenon; one cannot play a piece of music all at once or hear a piece of music all at once. Clearly, perceiving temporal order is fundamental to perceiving music, but it does not necessarily follow that temporal order determines the way in which the composition is stored in long-term memory. When recognition of a composition involves retrieval from long-term memory, is this retrieval process dictated by the temporal order in which the composition was perceived and learned? If we hear an excerpt from a well learned composition, does the time it takes to recognize the excerpt depend upon its temporal position in the composition? If the excerpt is near the end of the composition, does one have to scan long-term memory, so to speak, sequentially, in order to recognize the excerpt? If so, it should take longer to recognize an excerpt near the end of the composition than one that occurred near the beginning or middle. This study was undertaken to determine if the time required for a listener to recognize excerpts from musical compositions is dependent upon the temporal position of the excerpt within the composition (i.e., beginning, middle, or end).

Most previous research on memory for music has been focused on short-term memory for very short (two- to eight-note) sequences. In the typical recognition experiment, a subject listens to a short sequence of notes and then listens to the same or a different sequence. The subject's task usually is to respond "same" or "different" and sometimes to point out how the sequences differ (Cuddy & Cohen, 1976; Dowling, 1978). In the typical recall experiment, a subject listens to a short sequence of notes and then writes down or sings back the same sequence (Buttram, 1969; Williams, 1975). The primary reason for the use of short-term memory designs seems to be experimental control over the musical stimulus. Deutsch (1972, 1978) and others have taken the position that an understanding of how the basic components of the musical stimulus (presumably single notes or intervals) are remembered and recognized is critical to understanding more complex tonal relationships. Using very short sequences of notes and controlling for harmonic, rhythmic, and dynamic influences, researchers manipulate the pitches of individual notes or the sizes of intervals under the assumption that their research will yield basic information about memory for music.

Although the ability to differentiate and identify single notes or intervals is undoubtedly important to music perception, it has not been demonstrated that the kind of processing that occurs in short-term memory also occurs in long-term memory or that the conclusions derived from short-term memory research will lead to a better understanding of how individuals remember music. Whatever difficulties individuals have in remembering the pitches of eight or nine notes or in remembering the sizes of intervals, most can easily remember many pieces of music. The difference, suggests Davies (1978), is that, unlike a short sequence of notes, a piece of music is not remembered as simply a series of discrete events. Rather, it is the shape and form of the piece, the perceptual organization imposed on tonal stimuli temporally ordered, that make a series of events musical. That listeners can remember a number of musical compositions is the basis for this study. At issue is whether the temporal order of musical phrases in a composition affects memory for and recognition of the phrases. Clearly, perceiving temporal order is fundamental to perceiving music, but it does not necessarily follow that memory for and/or recognition of certain

phrases is determined by the order in which they occur in a piece of music. Short-term memory designs undeniably allow tight experimental control over the acoustic stimulus. Music, however, is not just a sequence of notes, and memory for music entails more than immediate (i.e., short-term) memory processes. Despite the difficulties inherent in the study of memory for actual pieces of music, an understanding of memory for music requires that such research be done.

Studies by Dowling (1973) and Ortmann (1926) are particularly relevant to the present research. In a series of experiments on recognition of two- to six-note sequences, Ortmann found that subjects made the fewest errors in recognizing changes in the first and last notes of a sequence. He also found primacy and recency effects when subjects were required to recognize sequences of intervals. Ortmann supposed that the first and last notes (or first and last intervals) served as perceptual anchors for the tonal sequence and were thus more salient. The extension of these findings to musical compositions suggests that serial position effects might be found in compositions as well, not for first and last notes or intervals but perhaps for entire measures or phrases.

Dowling's (1973) study was intended to determine if rhythmic grouping of stimulus items affects subjective chunking and memory storage of the items. It was not an investigation of serial position effects in music, and the "musical" stimuli were more analogous to the words and nonsense syllables used in verbal learning studies than to excerpts from musical compositions. Using series of rhythmically separated, five-note sequences, Dowling found the J-shaped serial position curve typical of verbal learning studies. In verbal learning studies, and presumably in Dowling's study as well, the recovery effect seems to be solely a product of short-term memory, and it disappears if recall is delayed (e.g., Glanzer & Cunitz, 1966). If results with delayed recognition of musical compositions were similar to those found in verbal learning studies, one might expect a primacy effect but no recency effect. If, as Ortmann (1926, 1933) suggested, however, last notes and intervals are more salient because they serve as perceptual anchors, then both primacy and recency effects in recognition of excerpts from musical compositions would be predicted.

In Ortmann's (1926, 1933) studies and in a study by Jeffries (1972) on rating intervals in terms of familiarity and frequency of usage, intersubject reliability was high. When only short sequences of notes are used and rhythmic and harmonic influences are minimal, memory demands on the subject are also minimal. As other components are added, the complexity of the musical stimulus may increase, and it is unlikely that a listener's attention is directed equally to all aspects of the music. Prince's (1972) paradigm for music listening research suggests several characteristics of the listener that might conceivably influence the number of musical events a subject is able to grasp at any one time and the kinds of events for which the subject listens. These characteristics include the listener's musical training and experience, music memory and musical ability, personality, and socially derived attitudes toward music. It seems likely that increasing the complexity of the musical stimulus will increase the variability among subjects, and it is not at all clear that different subjects will respond the same way to any one composition or that individual subjects will respond the same way to different compositions. For these reasons, a mixed-model design was used here.

METHOD

Subjects

Eight graduate students from the University of Georgia, ranging in age from 23 to 35 years, volunteered to participate in this study. Only one subject had not received formal training in music. The other subjects had 2-8 years of musical training, and six of these had played at least one instrument a minimum of 6 years. Two subjects did not complete the experiment. Experimental conditions were the same for all subjects.

Experimental Design

Subjects were run individually in daily sessions of 20-45 min, depending upon the stage of the experiment. In an initial screening session, subjects listened to each of the musical compositions used in this study to insure that the compositions were not familiar ones. None of the subjects recognized the compositions. Subjects learned compositions one at a time, and tests for temporal order effects in recognition of excerpts were conducted after each composition was learned.

Materials

Learning phase. The musical compositions were four tape-recorded Debussy études for piano. These études, labeled Compositions 1-4, were, in order, exercises in octaves, thirds, scales, and repeated notes. Each was approximately 3 min long, with distinct (nonrepetitious) beginning, middle, and end sections. Four versions of each composition, with certain sections modified by repetition, deletion, or addition of measures, were also recorded. Each modified version contained one to three altered sections of one to four measures each. Modified versions differed in length from the original composition by no more than ±5 sec.

The modified versions were to be used as a measure of learning, so it was important that subjects were not cued to the modifications by nonlearning factors such as poor splicing. As a control, nonsubject volunteers, unfamiliar with the études, listened to the recordings with instructions to report disruptions in the flow of music. The tapes were corrected until no modifications were noticed.

Testing phase. Also recorded on tape were 15 excerpts from each of the four études. Each excerpt was approximately 10 sec long, and there were five each from the beginning, middle, and end sections of each étude. The occurrence of natural breaks in a composition set limits to the selection of excerpts within sections, and wherever possible, excerpts began after natural rests or whole notes. Also recorded were 45 excerpts from Debussy études other than the four études used in this study. Additional student volunteers listened to the excerpts and rated the clarity and fullness of the first note or chord in each excerpt. Excerpts were retaped if a rater reported that a note sounded as if it had been cut in half.

All musical excerpts were recorded on one track of the tape. For each presentation of an excerpt, the temporal sequence of events was: warning tone, 1 sec; pause, 1 sec; excerpt, 10 sec. A 1,000-Hz tone was recorded on the second track. This tone automated an Iconix electronic clock that began counting simultaneously with the excerpt beginning. This clock was

stopped by the subject's depression of the response key. Response times were measured in milliseconds to an accuracy of ±.1 msec.

Two sets of 3 x 5 in. stimulus identification cards were used during the final testing phase to inform subjects of the correct excerpt label.

Apparatus

During all sessions, the subject sat in a sound-insulated chamber and listened to the musical stimuli through TDH-49 headphones. The sound-pressure level was adjusted to suit the subject each time a new composition was presented. A response key- was positioned directly in front of the subject for use during the testing phase of the experiment.

Procedure

Learning phase. Each learning session began with two successive presentations of the composition to be learned. This was followed by a test for learning that consisted of presentations of one or more modified versions of the composition and/or the unmodified version. The subjects' task was to listen carefully as a test tape was played and to indicate on a form provided them any modifications they detected (repetitions, deletions, or additions) or to indicate "none" if there were no changes. Subjects were told that each test tape contained zero to three changes. They were not told how many test tapes were made, nor were they corrected when they made errors in identifying modifications.

If the subject correctly identified the changes in the first test tape, the procedure was repeated with a second test tape, and if again correct, with a third test tape. If the subject could not identify the changes in a test tape, no additional test tapes were presented on that day. Test Tapes 1-5 (one of which was the original composition) were presented in the same order to all subjects. A subject was considered to have learned a composition when all changes in three test tapes, played within a session, were correctly identified. At the end of each learning session, the original composition was presented one more time.

The étude based on repeated notes was the first composition all subjects learned. This was considered a practice composition and was used to familiarize the subjects with the experimental procedures. The data from this composition were not used in subsequent analyses, and no identification tests were done on excerpts from this composition. The order in which subjects learned the other three compositions was random for each subject. For all subjects, the compositions were referred to as 1, 2, and 3, according to the order in which they were learned.

Recognition testing. Testing for recognition of excerpts took place on the day following criterion performance on the learning task. This procedure was used to determine (1) the time required by subjects to recognize excerpts taken from the beginning, middle, and end sections of a composition and (2) error rates for the identification of excerpts by temporal position (i.e., beginning, middle, or end). The recognition test consisted of two presentations of a series of 30 randomly ordered excerpts, half of which were taken from the étude the subject had just learned and half from the other études. The subject's task was to press the response key as quickly as possible after deciding whether each excerpt was or was not from the experimental composition. Subjects then recorded "yes" or "no" on a form and also indicated whether each recognized excerpt occurred at or near the beginning, in the middle, or at or near the end of the composition. The experimenter recorded the response time and reset the electronic clock after each excerpt. Approximately 15 sec elapsed between presentations of excerpts.

RESULTS

Learning Phase

The mean number of days required to learn Compositions 1, 2, and 3 were 5.2, 5.0, and 6.3, respectively. A one-way analysis of variance with repeated measures indicated no significant differences among these means [F(2,10) = .76, p > .05].

Recognition Testing

For Compositions 1, 2, and 3, respectively, the total number of recognition errors were four, four, and three. All errors represent failures to recognize excerpts. Median recognition times and interquartile ranges were calculated for each of the three compositions from the data obtained on the second series of trials (Table 1). For the single case in which a recognition error was made on that trial, the correct recognition time from the first trial was used.

Extremely skewed distributions precluded the use of parametric tests for determining if differences among recognition times varied as a function of excerpt position. Therefore, the nonparametric Friedman test was used. No significant differences among excerpts were found for Compositions 1 or 2 [$\chi_r^2(14) = 12.99$ and 20.72, respectively; p > .05]. Excerpts differed significantly in Composition 3 [$\chi_r^2(14) = 41.12$, p < .001]. Since there was no concentration of large or small interquartile ranges in any particular section of the composition (i.e., beginning, middle, or end), no further analyses were done.

For each composition, Kendall's coefficient of concordance was used to determine the degree to which ranked recognition times were the same among subjects. For Compositions 1, 2, and 3, respectively, W = .25, .15, and .49. As these values indicate, excerpts that were quickly recognized by some subjects were not necessarily quickly recognized by others.

Subject Questionnaire

In response to the question concerning difficulties in distinguishing between sections of compositions, no

Table 1
Median Recognition Times and Semiinterquartile Ranges (in Milliseconds)

Excerpt	Composition 1		Composition 2		Composition 3	
	Median	Range	Median	Range	Median	Range
Beginning						
1	789	313	716	353	1228	610
2	752	255	920	261	897	228
3	821	376	720	261	574	148
4	634	537	868	301	685	191
5	739	425	979	350	774	258
Middle						
6	748	164	738	450	781	430
7	939	260	597	181	692	180
8	835	302	729	244	636	321
9	569	212	600	522	805	98
10	889	373	729	148	910	456
End						
11	1028	550	807	297	1177	679
12	881	453	670	417	931	389
13	916	604	821	271	590	138
14	662	180	796	163	706	456
15	676	250	544	54	845	1165

two subjects mentioned the same pair of sections. Only two sections, the beginnings of Compositions 2 and 3, were not mentioned by any subject. Subjects also differed in response to the question concerning distinctiveness of sections. Three of the six subjects considered the beginning of Composition 3 distinctive, but no other sections were called distinctive by more than two subjects. All subjects recognized the scale-like quality of Composition 3, two subjects realized that Composition 2 was a study in thirds, and two subjects believed that Composition 1 was a study in octaves.

DISCUSSION

The results of this study indicate that whether a sequence of notes is at the beginning, middle, or end of a composition does not in itself determine recognition time. The temporal order in which we must hear a composition does not dictate how it will be remembered. It does not appear that we have to scan long-term memory sequentially in order to recognize excerpts from different temporal parts of a musical composition.

"Knowing" a piece of music involves at least three factors. One is recognizing the piece, that is, distinguishing it from other pieces. Another is identifying it: giving it a unique name. The third is recognizing temporal order: Has the piece just begun, is it ending, or is it somewhere in the middle? None of the subjects had any difficulty recognizing the pieces. During the experiment, one of the subjects mentioned that he often knew the position of an excerpt before he knew from which composition it was taken. The other subjects were asked if this was true for them, and five of the six agreed. Their estimates of the percentage of trials on which this occurred ranged from 10 to 90. The subjects' remarks were supported by the data.

The results throughout this study show large individual differences. Much of the variability seems to be because excerpts that are relatively easy for one subject to recognize are not easy for others. This was also the case for identification times. As anticipated, subjects seemed to be responding to different aspects of the musical stimulus. That these represent real differences and not simply "noise" is supported by subjects' comments during the course of the experiment and by their answers to questions at the end of the study.

Some degree of concordance among subjects was evident in data from Composition 3. All subjects recognized the scale-like quality of that composition, and it was the only composition in which there was some amount of agreement in excerpt recognition as indicated by the Kendall test. It seems likely that subjects used the scale-like quality as a key to identifying the composition. Because all were attending to the same aspect of that composition, there was more conformity among subjects.

Many years ago, Wundt (1900) pointed out that in the generation of sentences, a simultaneous cognition is transformed

into a sequential structure. If we see Johnny running, it is incorrect to assume that because we say, "I see Johnny running," we have a sequence of cognitions: that we first see Johnny and then we see some running. Instead, "seeing Johnny running" is a single cognitive structure that is transformed into a sequential linguistic structure. In the case of music, what we found suggests that sequential cognitions are transformed into a simultaneous cognitive structure, any aspects of which can be retrieved from memory storage in the same length of time. Although we have to hear the beginning of a composition before we hear the end, this temporal order in perception is not determinative of the cognitive structure retrieved from memory.

REFERENCES

BUTTRAM, J. B. The influence of selected factors on interval identification. *Journal of Research in Music Education*, 1969, **17**, 309-315.

CUDDY, L. L., & COHEN, A. J. Recognition of transposed melodic sequences. *Quarterly Journal of Experimental Psychology*, 1976, **28**, 255-270.

DAVIES, J. B. *The psychology of music.* Stanford, Calif: Stanford University Press, 1978.

DEUTSCH, D. Music and memory. *Psychology Today*, December 1972, pp. 87-89; 119.

DEUTSCH, D. The psychology of music. In E. C. Carterette & M. P. Friedman (Eds.), *Handbook of perception* (Vol. 10). New York: Academic Press, 1978.

DOWLING, W. J. Rhythmic groups and subjective chunks in memory for melodies. *Perception & Psychophysics*, 1973, **14**, 37-40.

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

GLANZER, M., & CUNITZ, A. R. Two storage mechanisms in free recall. *Journal of Verbal Learning and Verbal Behavior*, 1966, **5**, 351-360.

JEFFRIES, T. B. Familiarity-frequency ratings of melodic intervals. *Journal of Research in Music Education*, 1972, **20**, 391-396.

ORTMANN, O. On the melodic relativity of tones. *Psychological Monographs*, 1926, **35**, 1-47.

ORTMANN, O. Some tonal determinants of melodic memory. *Journal of Educational Psychology*, 1933, **24**, 457-467.

PRINCE, W. F. A paradigm for research on music listening. *Journal of Research in Music Education*, 1972, **20**, 445-455.

WILLIAMS, D. B. Short-term retention of pitch sequence. *Journal of Research in Music Education*, 1975, **23**, 53-66.

WUNDT, W. *Völkerpsychologie* (Vol. 1, Book 2) (*Die Sprache*). Leipzig: Engelmann, 1900.