Perception & Psychophysics
1988, 43 (4), 389-394

Auditory perceptual learning of tonal patterns

MARJORIE R. LEEK and CHARLES S. WATSON
The Boys Town National Institute for Communication Disorders in Children, Omaha, Nebraska

Characteristics of auditory perceptual learning were investigated by monitoring improvements in the identification of tonal patterns ranging in length from 135 to 540 msec in total duration. Increasing amounts of temporal complexity were imposed by the combination of elementary 3-tone segments into sequences of two, three, or four segments, thus creating patterns of 3 to 12 tones that varied in frequency and duration. Four of 5 listeners approached asymptotic identification accuracy near 90% for the longest patterns (12 tones) after about 20 h of experience with these stimuli over several weeks of daily practice. Identification was most accurate for the initial and final portions of each pattern and poorest for the middle sections. Large individual differences among subjects were apparent in the order in which the patterns were learned and in the difficulty of the various patterns for each listener. These findings suggest differences in learning strategies or differences in the focusing of auditory attention.

Investigations of listeners' perceptions of various auditory stimuli have typically measured the end result of perceptual learning. Most psychoacousticians ignore the training phases of their studies, simply reporting that their data were obtained from "well-practiced listeners." There is usually at least tacit acknowledgment, however, that a listener's performance changes during an experiment. Data collected after some period of training, which are characterized by improved levels of performance and increased consistency and stability, are assumed to more accurately reflect listeners' true sensory capabilities.

The goal of the present study was to examine the process of learning to recognize temporally complex stimuli. Sequences of pure tones were constructed with some systematic constraints on the allowable order of individual tonal elements. Improvements in the ability to identify these patterns were monitored over a period of several weeks of listening. In addition to providing information concerning the possible limits on learning to recognize a small set of tonal patterns, this experiment was designed to determine the time course of improvements in stimulus identification and to gain some insight into the qualitative aspects of auditory perceptual learning.

Descriptions of identification learning are characterized by long training times and significant differences among individual subjects. In an early study of the time course of auditory learning, Bryan and Harter (1899) reported continued improvement in the accuracy of Morse code reception over a training period lasting more than 48 weeks. Webster, Carpenter, and Woodhead (1968a, 1968b) trained listeners to identify nine harmonic complexes over a period of 2 weeks. Although the listeners could readily discriminate the complexes, even before training, the ability to identify the sounds remained poor, reaching accuracy levels of only 30% to 40%, and performance was quite variable across subjects.

In 1969, Warren, Obusek, Farmer, and Warren reported that naive listeners were unable to determine the order of occurrence of a repeating cycle of four sounds unless each component was more than 200 msec long. This result was unexpected, inasmuch as the order of sound sequences is routinely identified at much shorter durations in everyday speech. Subsequent studies, motivated by this unusual finding, have shown that subjects with extensive listening experience with the sound sequences can perform the identifications with much shorter component durations. Neisser and Hirst (1974) used the same sound sequences that had been used in the Warren et al. study (a combination of pure tones and noises), and found substantial decreases in the component duration required by the listeners after training for 28 h spread over 7 weeks. In this study, the listeners responded with a series of four digits, each corresponding to a component of the sequence (e.g., "2,3,4,1").

Nickerson and Freeman (1974) suggested that part of the difficulty of this task was associated with the need to construct a response "string" after each stimulus presentation. Thus, although the actual perception of temporal order might be possible at much shorter component durations, the designated sequential response code might require a relatively inefficient and slow decoding of the stimulus sequence. Nickerson and Freeman asked their subjects to label each sequence, made up of a different order of four component sounds, with only one digit (1

The research reported here was supported by Grant NS 14637 from the National Institute for Neurological and Communicative Disorders and Stroke (NINCDS) to the Boys Town Institute for Communication Disorders in Children, Omaha, Nebraska, and by an NIH postdoctoral fellowship to the first author. Portions of this research were presented at the 104th meeting of the Acoustical Society of America, Orlando, 1982. Preparation of this article was partially supported by a grant from AFOSR to Indiana University. M. R. Leek's current address is Department of Communication Disorders, University of Minnesota, 164 Pillsbury Dr. S. E., Minneapolis, MN 55455. C. S. Watson's current address is Department of Speech and Hearing Sciences, Indiana University, Bloomington, IN 47405.

to 6), rather than to construct a response by a concatenation of component labels. Unfortunately, they trained their 6 subjects for only 2 days, and they were unable to report much improvement in identification accuracy for components with 200-msec durations. However, they also reported the effects of extensive training for 1 subject, who, after several months of practice, achieved 90% or better accuracy for even extremely brief components. They concluded that, for components of relatively long duration, an identification response was generated by a reconstruction of the order of the components in the sequence; the response to stimuli of shorter duration was mediated either by qualitative characteristics of the entire sequence or by the order of subsets of the components.

The present paper describes some aspects of the learning process for several sequences of tones. The sequences were constructed of from 3 to 12 tones that varied in frequency and duration. At the simplest level of temporal complexity, 4 different 3-tone patterns were presented to the listeners for identification. This proved to be a relatively simple learning task, and so increasing levels of complexity were imposed by the concatenation of the 4 short patterns into sequences of two, three, and four segments. The maximum level of complexity tested was the set of 24 tonal patterns constructed by ordering the four three-tone sequences in all possible ways. Thus, for the longest patterns, the 4 components making up these sequences varied in frequency over time, unlike the simple steady-state tones and noises used as components in previous studies. These stimuli also differed from those used in earlier studies on temporal order identification in that they did not continuously repeat; rather, they were constructed as individual patterns that required an identification response after every presentation.

Tonal sequences were chosen for three reasons. First, they were sufficiently unlike any familiar sounds heard in everyday life for us to expect to be able to measure the entire course of perceptual learning. Second, the use of tones as the elementary components of patterns allowed us to impose some similarities across stimuli so that differences along acoustic dimensions, such as durations and bandwidths, would play a minimal role in the learning process. Concatenation of the four elementary three-tone segments in different orders allowed us to produce different patterns but still be assured that the physical characteristics of each were essentially the same. Last, previous studies of pattern discriminability (e.g., Watson, Wroton, Kelly, and Benbasset, 1975) allowed us to choose a set of patterns that could be distinguished readily in pairwise comparisons but did not mix stimulus types, such as steady-state noises and tones.

METHOD

Subjects

Five audiometrically normal listeners participated in the study. They ranged in age from 20 to 35 years. The length of their participation in the study depended on the speed with which they learned the sequences and their willingness to continue. All listeners participated in the identifications of the single segments and in the four-segment concatenation tasks, but not all performed the two- and three-segment concatenations. The minimum time spent by any listener on this series of identifications was about 6 weeks.

Stimuli

Auditory patterns consisting of sequences of pure tones were selected as the catalog to be learned. All frequencies were within a range of 300 to 3000 Hz, and the total duration of the patterns ranged from 135 msec for the elementary 3-tone segments to a maximum of 540 msec for the 12-tone sequences.

Figure 1 illustrates the four elementary triplets of tones that were combined to create the catalog of patterns. The assigned response labels, digits 1 to 4, are shown below each segment. The duration of each individual tone was between 25 and 65 msec, with the constraint that each segment was 135 msec in length. These sets of tones were designed to be easily pairwise discriminable on the basis of their mean frequencies, their frequency range, the temporal patterns of their component durations, and the form of their pitch contours.

Pure tones were produced by a voltage-controlled oscillator gated on and off through an electronic switch with a 2.5-msec rise–fall time. There were no silent intervals between the tones, either within the elementary segments or when the segments were combined into longer patterns. The sequences were presented monaurally over earphones at 75 dB SPL. The subjects were tested either individually or in pairs in a sound-treated booth. The presentation of the stimuli and the collection of responses were controlled by computer.

Procedures

Each listener was seated in front of a computer terminal. Immediately after a warning prompt appeared on the CRT ("Listen!"), a stimulus was presented. The listener's task was to enter on the keyboard a response consisting of one to four digits that indicated the order in which the segments were heard. After a response was entered, the CRT displayed the correct order of the segments and the word *correct* or *incorrect*. The correct sequence of segment labels remained on the CRT until the listener called for the next trial presentation, thus allowing the listener to study the correct answer as long as desired. Listeners were encouraged to work as quickly as was comfortable, but accuracy was stressed over speed.

The first group of stimuli the listeners heard were the four three-tone segments; they were instructed to enter one digit in response to each stimulus. The four individual segments were presented at

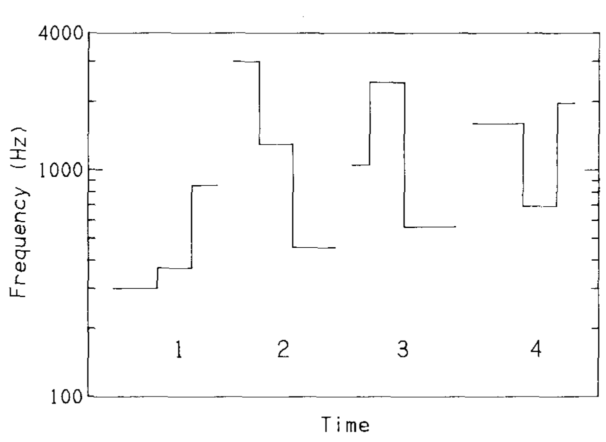

Figure 1. Four tonal segments, each made up of 3 tones varying in frequency and duration. Each segment was 135 msec in length. Combinations of these segments created patterns containing from 6 to 12 tones.

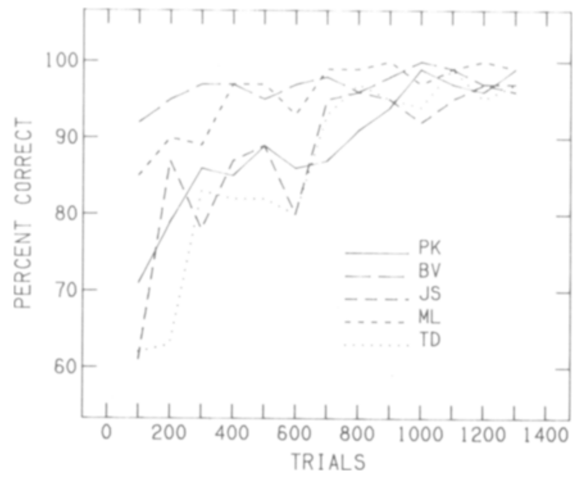

Figure 2. Mean percent correct identifications across four individual segments for 5 listeners, shown as a function of trial presentations.

random, in 100-trial blocks. Immediately following each response, the correct answer was shown on the CRT as feedback.

Listening to single segments continued until all subjects were performing at better than 90% identification accuracy for each segment. Identification for longer sequences was then measured by presenting combinations of the four elementary segments; a response of two, three, or four digits was required to reflect the listener's identification of the order of occurrence of the segments. The pattern length conditions were varied in order (two-, three-, or four-segment combinations), and all trials of one length were completed before a condition with a longer pattern length was begun. Four of the listeners identified pairwise combinations of the segments. Three of these listeners then participated in the three-segment combinations, and then all 5 listeners responded to the longest patterns, consisting of 12 tones (all combinations of four 3-tone segments).

For all pattern length conditions, the stimuli were presented in blocks of 96 trials, with equal numbers of presentations for each possible combination randomized within the blocks.

RESULTS

Figure 2 shows the learning curves for the identification of the elementary segments by the 5 listeners. These are means across the four different segments for each listener. After about 800 trials (200 trials per segment; a total of about 2 h of listening time), all listeners were performing at better than 90% accuracy for all four segments. Note that although performance at the beginning ranges from about 60% to about 90% correct for the 5 listeners, all listeners were performing similarly after several blocks of trials.

The three panels of Figure 3 show learning curves for each listener on the two-, three-, and four-segment patterns. For 4 of the 5 listeners, asymptotic performance above 80% correct was approached on each of the identification tasks after 40–60 training blocks (96 trials per block). Listener J.S. showed a similar course of improvement for the two-segment task, but demonstrated little learning after the first 10 blocks on the three- and four-segment tasks. On the latter tasks, J.S. never consistently exceeded 40% correct, despite intensive practice.

The right panel in Figure 3 displays the learning curves for the longest patterns. The best performance is shown by 2 of the listeners who participated in the shorter pattern conditions (P.K. and B.V.). Slower learning was demonstrated by the other 2 listeners: 1 listener (M.L.) had heard no previous combinations, and the other (T.D.) participated only on the two-segment combinations. The similarity of these two curves suggests that the extra train-

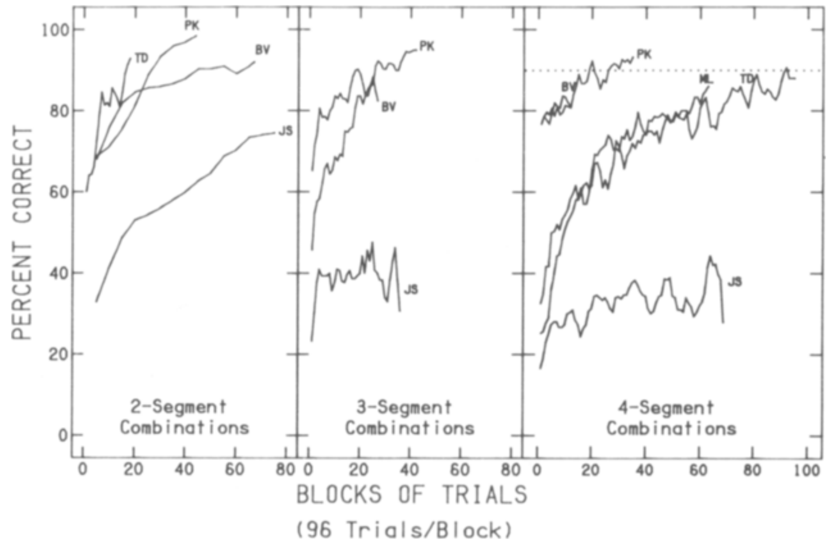

Figure 3. Percent correct identifications of patterns by each listener as a function of the number of blocks of trials. The left panel represents performance on the two-segment combinations; the middle panel shows performance for three-segment combinations; and the right panel shows performance for four-segment patterns.

ing provided by the two-segment combinations was not a major factor in determining the final level of identification performance.

Transfer of Learning

Comparisons across the panels in Figure 3 can provide some insight into the degree of transfer of learning that may have occurred as the listeners advanced to longer patterns. First, it appears that the two-segment trials were of little benefit in preparation for the three- or four-segment patterns. All listeners' performances dropped considerably at the beginning of the three-segment trials. T.D., who did not listen to the three-segment trials, also showed the performance drop at the beginning of the four-segment identifications.

The three-segment trials, however, may have provided valuable experience for at least 2 of the listeners. Note that, after three-segment training, B.V. and P.K. start at considerably higher performance on four-segment listening than do listeners without the lower level listening experience.

Time Course of Learning

The learning of these auditory patterns is reflected in decreased response latencies as well as in increases in accuracy. At the beginning of the long pattern identifications, listeners generally required over 15 min to make 96 responses, or an average of 9 sec per trial. By the end of their listening experience, they were performing at more than double that speed. It should be noted, however, that even an average of 3-4 sec per trial is a considerable amount of time to identify a single pattern, and suggests a studied construction of the response sequence, as suggested by Nickerson and Freeman (1974).

Estimates of the total listening experience required to learn to identify this catalog of patterns with a high degree of accuracy may be developed on the basis of the learning curves shown in Figure 3 and the time required to make each identification. In view of the clearly different course of performance demonstrated by listener J.S., and the decreased final accuracy shown, these data are not included in this or subsequent analyses of learning characteristics. The problem of individual differences in learning, of which this is a classic example, will be taken up in the discussion.

As noted earlier, response times decreased over the course of the experiment, so an average of 10 min per block of 96 trials is assumed. Projections were calculated on the basis of Figure 3, by first estimating a probable trials-to-asymptotic performance for each listener on the four-segment tasks, and adding that to the amount of practice each listener experienced with the shorter concatenations.

Listener P.K. reached 90% accuracy within about 30 blocks of training on the set of long patterns; this took about 5 h of listening. The same listening time might be estimated for Listener B.V. on the basis of the similar learning curve. The two middle curves suggest that M.L. and T.D. would approach asymptotes after approximately 80 to 100 blocks of trials, or about 13 to 17 h of practice.

For these 4 listeners, the total learning times are quite similar, ranging from 17 to 20 h of training (see Table 1), even though the exposures to two-, three-, and four-segment stimuli varied across subjects. If listening to the shorter stimuli had provided no benefit in identification learning for the longest patterns, longer total exposures would be expected for the listeners who also identified the shorter patterns. The similar projected learning times, based on total exposure to the stimuli, provide further support that there was some transfer of learning from the shorter duration patterns to the longer ones.

Details of the Learning Process

Insight into the mechanisms of perceptual learning may be gained by examining some of the characteristics of the 4-segment patterns and the listeners' responses to them. There were no obvious acoustic characteristics which, a priori, would make any of these 24 patterns more or less difficult than any others. Indeed, a pattern-by-pattern analysis of the data showed that there were no patterns that were difficult for all the listeners, or, conversely, were learned quickly by all the listeners. To determine whether the listeners tended to learn the 24 patterns in a similar order, Kendall's rank order correlations (tau) for the order in which the patterns were learned by each listener were calculated. A significant positive correlation for any of the six pairs of listeners would have suggested a similar order of pattern learning for those 2 listeners. However, none of the correlations approached statistical significance ($p > .1$ for all listener pairs), in-

Table 1
Estimated Learning Times

	Number of Trial Blocks		
Listener	2- and 3-Segment Blocks (Completed)	4-Segment Blocks (Estimated)	Total Learning Time (Hours)
P.K.	80	30	18.3
B.V.	85	30	19.2
T.D.	19	100	19.8
M.L.	0	100	16.7

Note—Number of blocks required for 90% correct performance on the complete 4-segment patterns is projected from the curves in Figure 3. Total learning times are estimated at 10 min per 96-trial block for the two-, three-, and four-segment training combined.

dicating a considerable degree of idiosyncracy in performance across these individuals.

To determine the effect of the placement of a segment within a sequence on identification accuracy, or whether any segment in the sequences was identified more accurately than the others, a partial scoring of the data was undertaken for trials early in the listening experience and toward the end. The values in Figure 4 were calculated by scoring any response as correct if the segment in question was identified correctly, whether or not the remaining segments were identified in the correct order. (This contrasts with earlier scoring, in which responses were correct only if *all* segments in a stimulus were reported in correct order.)

Figure 4 shows the mean percent correct identification of each segment scored individually over the first 20 presentations of each pattern (labeled "Beginning trials") and the last 20 presentations ("Ending trials") in the three- or four-segment conditions for the 4 listeners. Figure 5 shows a similar analysis by position within the pattern. (The "Middle" category combines Positions 2 and 3.)

Performance early and late in training with these patterns was similar across segments (Figure 4) and across position (Figure 5). At both times, Segment 1 was most accurately identified, followed in order by 4, 2, and 3. The amount of improvement with experience was similar for each segment: mean accuracy improved by 28% for Segments 1 and 2 and by 31% for Segments 3 and 4 from the early trials to those later in training, indicating that learning was distributed uniformly over the four segments of the patterns.

Figure 5 shows that, both early and late in training, the best identification occurred for the last part of the patterns (a "recency" effect), followed by the first part of the patterns ("primacy"). This finding is in agreement with other studies of the sequential ordering of sounds (e.g., Divenyi & Hirsh, 1978), although at least one earlier study of pattern discrimination reported recency but not primacy effects (Watson et al., 1975).

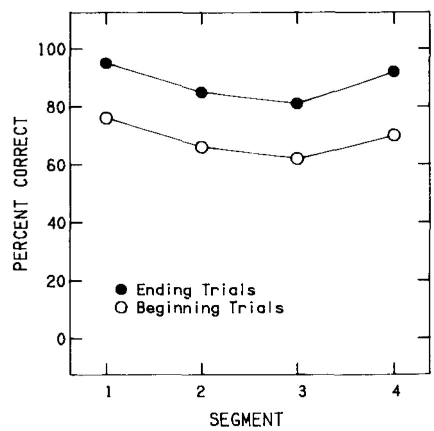

Figure 4. Mean identification scores for each segment within the 12-tone patterns over the first 20 trials per pattern ("Beginning trials") and over the last 20 trials ("Ending trials").

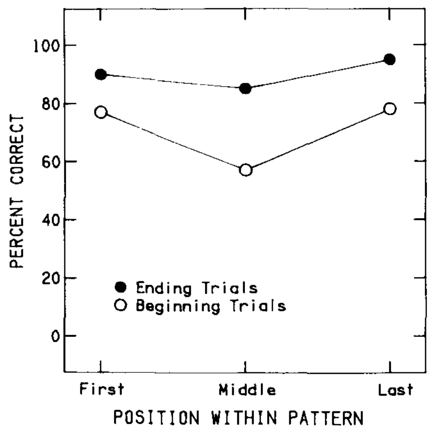

Figure 5. Mean identification scores for segments in the first, middle, and last positions in the 12-tone patterns over the first 20 trials per pattern ("Beginning trials") and the last 20 trials ("Ending trials").

DISCUSSION

This experiment investigated the largest catalog of nonspeech sounds for which systematic learning data have been reported. The listeners in this study approached asymptotic learning performance after about 20 h of training. Several aspects of the experimental procedure may have influenced auditory learning in a positive manner. The response—typing digits on a keyboard—was simple. Experimental sessions were typically held on a daily basis so that the listeners practiced at regular periodic intervals. The listeners were encouraged to take as much time as they needed to study the feedback, and to silently rehearse the pattern they had just heard. The trials were presented on the listeners' demand, perhaps reducing the incidence of careless misidentifications or lapses of attention. In addition, although the most temporally complex patterns were combinations of 12 separate tones differing in frequency and duration, the listeners were trained to cluster them into four groups of 3 tones each, so that a strategy of ordering four segments rather than 12 tones could be used.

There were some interesting similarities in performance across 4 of the 5 listeners. The total experimental time to their final high levels of accuracy in identifying the long patterns ranged from 17 to 20 h, even though different proportions of that total time were devoted to each level of pattern duration or complexity across subjects. In addition, the individual learning curves for these listeners pointed to a similar final identification accuracy of about 90% for all. All listeners showed recency and primacy effects in their abilities to identify portions of the long patterns.

However, as is characteristic of studies that require complex responses to complex sounds, there were significant individual differences in the performance of these listeners. The idiosyncratic nature of the interaction between subjects and stimulus difficulty, which was reflected in the different orders in which the long patterns were learned by each subject, suggests that each person was focusing attention in a unique way. Such individual differences are typical of many tasks involving complex auditory signals (e.g., Leek and Watson, 1984). Also not unusual in complex sound experiments is the puzzling performance of some listeners, which is quite dissimilar to that of their colleagues in the listening booth. Our atypical listener was like the other listeners in all relevant respects, displaying similar performance on such simple auditory tasks as detection thresholds and individual segment labeling. His identification performance on the combinations of segments, even the most complex ones, was not characterized by a complete lack of learning, in that all combinations were identified at levels well above chance. Nevertheless, as the complexity of the task increased, the performance of this listener was consistently different from that of the other listeners.

Neisser and Hirst (1974) also reported large intersubject differences in temporal ordering of sound sequences. They suggested that the listeners' individual abilities and strategies, which were present at the beginning of the experiment, were as great a determinant of performance as were the experimental manipulations of the physical stimuli.

Although Neisser and Hirst (1974) were investigating the lower limits of component duration for correct temporal ordering of sequences, their study may be viewed as a simpler version of the present task. In both investigations, four sounds were concatenated in all possible orders, and the listeners were asked to report the order of occurrence. However, Neisser and Hirst's four elementary components were all simple, steady-state sounds, such as noises and single-frequency tones. The authors hypothesized that the listeners performed the identification by constructing a response sequence after the stimulus presentation on the basis of more-or-less complete "analog" representations of the four sounds, and then produced the responses required by the experiment. Successful implementation of such a strategy requires the subject to hold some representation of the four sounds in a memory store until a response sequence can be designed (e.g., "hiss, low tone, buzz, high tone").

In the present study, the designation of the pattern labels as combinations of previously learned segment labels probably encouraged a similar strategy. The resulting heavy load on auditory memory may have been partially responsible for the differences in performance among the listeners. In fact, some recent evidence suggests that a major source of variance among listeners may be the informational capacity of the early stages of auditory memory. Watson and Foyle (1983) reported that a limit on tonal pattern discrimination performance seems to be the number of elements in a pattern, independent of their duration or of the duration of the total pattern. Listeners frequently perform "normally" in terms of single-tone detection or discrimination, yet show large individual differences when a tonal sequence contains more than five to seven tones.

The results of this study may be useful in providing a starting point for further investigation of the time course of auditory identification learning for sets of nonspeech stimuli of comparable size and pattern length. The auditory code with which we have the most learning experience is, of course, speech, and we may take some hints from the structure of ongoing speech which may prove fruitful in increasing the rate or accuracy of perceptual learning of nonspeech stimuli. For example, an increase in the learning of a set of 24 0.5-sec tonal patterns might be evidenced by inserting silent intervals between segments, by including some harmonic structure, or by providing smooth transitions between large frequency steps within the patterns. The inclusion of these three properties of speech (as well as others) into tonal patterns may result in improvements in auditory learning over that demonstrated in this study, and might thereby contribute to our knowledge of the importance of such characteristics in learning speech.

REFERENCES

BRYAN, W. L., & HARTER, N. (1899). Studies in the physiology and psychology of the telegraphic language: The acquisition of a hierarchy of habits. *Psychological Review*, **6**, 345-375.

DIVENYI, P. L., & HIRSH, I. J. (1978). Some figural properties of auditory patterns. *Journal of the Acoustical Society of America*, **64**, 1369-1385.

LEEK, M. R., & WATSON, C. S. (1984). Learning to detect auditory pattern components. *Journal of the Acoustical society of America*, **76**, 1037-1044.

NEISSER, U., & HIRST, W. (1974). Effect of practice on the identification of auditory sequences. *Perception & Psychophysics*, **15**, 391-398.

NICKERSON, R. S., & FREEMAN, B. (1974). Discrimination of the order of the components of repeating tone sequences: Effects of frequency separation and extensive practice. *Perception & Psychophysics*, **16**, 471-477.

WARREN, R. M., OBUSEK, C. J., FARMER, R. M., & WARREN, R. (1969). Auditory sequence: Confusion of patterns other than speech and music. *Science*, **164**, 586-587.

WATSON, C. S., & FOYLE, D. C. (1983). Temporal and capacity limitations of auditory memory. *Journal of the Acoustical Society of America*, **73**, S44.

WATSON, C. S., WROTON, H. W., KELLY, W. J., & BENBASSET, C. A. (1975). Factors in the discrimination of tonal patterns: I. Component frequency, temporal position, and silent intervals. *Journal of the Acoustical Society of America*, **57**, 1175-1185.

WEBSTER, J. C., CARPENTER, A., & WOODHEAD, M. M. (1968a). Identifying meaningless tonal complexes. *Journal of the Acoustical Society of America*, **44**, 606-609.

WEBSTER, J. C., CARPENTER, A., & WOODHEAD, M. M. (1968b). Identifying meaningless tonal complexes: II. *Journal of Auditory Research*, **8**, 251-260.

(Manuscript received January 12, 1987;
revision accepted for publication September 18, 1987.)