

# Identification of temporal order within auditory sequences\*

RICHARD M. WARREN and CHARLES J. OBUSEK  
University of Wisconsin-Milwaukee, Milwaukee, Wisconsin 53201

Unpracticed Ss reported the order of sounds in sequences consisting of either three or four successive items repeated over and over without pause. With unrelated sounds each lasting 200 msec, correct reports of order were at chance level for oral responses and for card-ordering responses (each card bearing the name of one sound). The sequences with four unrelated items were studied in greater detail, and the threshold for identification was found to be 670 msec with oral responses and 300 msec with card-ordering responses. When two related sounds (tones) were used in four-item sequences, correct card-ordering was possible at 200 msec per item when the tones were temporally contiguous, but was not possible at this duration when the tones were separated by nonrelated sounds. Some special rules governing auditory sequence identification were suggested, and implications for theories of auditory perception discussed.

Detection of the temporal order of sounds is an important perceptual ability; lacking this skill, speech (and music) could not have been developed by man. The average duration of phonemes in speech is about 80 msec (Efron, 1963), and the notes of melodic themes in musical compositions follow one another at intervals as low as 150 msec (Fraisse, 1963). It was with some surprise that we found recently that listeners could not identify the order of items in certain sequences containing three or four unrelated sounds (hisses, buzzes, tones) when each sound lasted 200 msec (Warren, 1968; Warren, Obusek, Farmer, & Warren, 1969). It should be noted that the figures cited for speech and music allow virtually error-free perception; durations of phonemes in "compressed speech" can be dropped to an average well below 50 msec before intelligibility is lost (Aaronson, 1967), and successive musical notes can be ordered down to almost 50 msec (Winckel, 1967).

The low temporal resolution for sequences of nonrelated sounds may seem at first to conflict with several earlier reports based upon pairs of such sounds (selected from electrically generated tones, buzzes, hisses, bings, and bongs), which indicated that it was possible to detect the order of the two sounds for separations down to 20 or 30 msec (see Fay, 1966, for a review of this literature). However, it seems that items occurring first and last in a series may be identified by

virtue of their presence at onset and termination of stimulation (see Warren & Warren, 1970). Thus, temporal localization need not be made relative to other components of the series (as in speech and music), but can be accomplished by recall of the first and/or last sound.

While some of our earlier experiments intentionally introduced onset and termination cues, we generally avoided such aids to identification by employing a series of four sounds repeated over and over without pause until after the S finished responding. He could name the sequence starting with whichever sound he wished (so there was no advantage in identifying the first sound heard)—but in order to be scored as correct, the next three sounds must have been named in correct sequence (the chance of guessing correctly is 1 in factorial 3, or one-sixth).

When we used four arbitrarily chosen sounds (selected from a number of hisses, buzzes, tones) each lasting 200 msec, the number of correct responses was not significantly greater than chance. When each 200-msec sound was replaced by a 200-msec digit-word, the correct order was stated accurately and with ease by every S. In addition, we found that the ability to order a series of four vowels (/i/, /A/, /I/, /u/) depended upon the closeness of the recorded sequences to normal utterances: it was poor for temporally contiguous 200-msec segments excised from extended statements of the vowels, improved when the duration of each phoneme was cut to 150 msec and a silent gap of 50 msec inserted between sounds, and still better for complete utterances of the separate vowels having normal onset and decay characteristics (Warren, 1968). Thomas, Hill, Carroll,

and Garcia (1970) also found that four recycled vowels (/i/, /ε/, /a/, /u/) were ordered more readily than nonspeech sounds, and Thomas, Cetti, and Chase (1971) confirmed that silent intervals separating contiguous vowels facilitated identification of temporal order.

The present series of experiments was designed to investigate further the ability to identify the temporal order of nonrelated sounds. We have continued our procedure of using separate groups of at least 30 Ss for each condition and limiting each S to a single judgment. This procedure has the inconvenience of requiring several hundreds of Ss for an extended study. However, there are considerations which led us to our decision to employ single rather than multiple judgments. In experiments requiring the temporal ordering of pairs of sounds, the reported values range from about 20 to 80 msec, apparently reflecting the specific amounts and type of training (see Malone, 1967). Some of our unpublished experiments with recycled sequences of four sounds have indicated also that practice, as would be anticipated, can bring about some improvement in performance. It seems that this improvement probably depends upon the specific type and amount of practice, and the precise relation of the practiced to the tested sequences. Since the present experiment measures identification of temporal order uninfluenced by prior training, the results reflect directly the actual (as distinguished from potential) ability in the population from which our Ss were drawn. The results also can be used as a base against which other experiments involving the effect of training may be compared.

## METHODS AND RESULTS Stimuli

When sequences consisted solely of items produced by electronic signal generators and filters, the signals were presented successively in a recycled continuous mode using Grason-Stadler Series 1200 programming equipment. Transition from one signal to the next was accomplished through an electronic switch set at a rise/decay time of 5 msec. For those sequences containing the speech sound /i/, separate recordings lasting several seconds were made of each item at a fixed level ( $\pm 1$  dB) on an Ampex PR10 recorder set at 15 ips. Segments of tape corresponding to the appropriate durations were cut from the separate recordings. These strips were spliced together in the desired order, with the final splice producing a closed loop. Since the loops were too fragile for continued use, a conventional reel of

\*This study was supported by grants from the National Science Foundation (GB-26459) and from the University of Wisconsin-Milwaukee Graduate School to the first author. We wish to acknowledge the many valuable suggestions made by Dr. Roslyn P. Warren, and to thank Mr. Gary L. Sherman and Mr. Richard M. Farmer for assistance with portions of the experimental work.

Table 1  
Effect of Changing Duration of Constituent Sounds [(Buzz, Tone, Hiss, /i/)<sub>n</sub>] Upon Sequence Identification for Two Response Procedures

Duration of Items (Msec)	Oral Response			Card-Ordering Response		
	Number Correct Out of 30	Response Time (Sec)		Number Correct Out of 30	Response Time (Sec)	
		Median	Q1, Q3		Median	Q1, Q3
200	7	36	23, 65	6	52.5	36, 76
300	9	26.5	15, 43	15	30	25, 54
450	6	24.5	15, 45	24	22.5	17, 32
670	16	25.5	15, 37	22	26	19, 35
670*	16	19.5	14, 36	—	—	—
1000	17	24	16, 23	18	24	20, 31

\*Replication of previous group using other Ss and a different E

tape lasting 3 min was rerecorded at 7½ ips.

The sequences were all repeated over and over without any pauses between successive sounds and were not terminated until S completed his response. The starting positions within the sequence varied randomly for Ss within each experimental group.

#### Subjects

A total of 550 Ss (15 separate groups of 30 Ss and 2 separate groups of 50 Ss) were recruited from undergraduate psychology courses. Each received either credit toward the course grade or \$1.00 for participation. Only a single judgment of one sequence was made by each S. None had participated previously in any experiment on sequence perception.

#### Procedure

Each S was tested individually in a large double-walled audiometric room (IAC Model 1204A) in which both he and the E were seated at a table. After reading the typewritten instructions for his particular experiment, S listened to a sequence of sounds delivered through a pair of matched TDH-49 headphones at a level of 80 dBC (re: 0.0002 microbars). The initial sound presented within each sequence was selected randomly for each S. They were instructed either to call out their answers or to arrange cards carrying the names of the stimuli in the appropriate order. E wrote down the responses and the time required for response before turning off the stimulus.

In each experiment, responses scored as correct could start with any one of the items in the repeated sequence.

*Experiment 1: Effect of duration of individual sounds upon oral responses.* Four nonrelated, readily distinguished, and easily identified sounds were used: (1) a buzz (40-Hz square wave); (2) a tone (1,000 Hz); (3) a hiss (2,000-Hz octave band noise); (4) the vowel /i/ (excised from an extended steady

statement). Separate stimulus tapes differed only in duration: each contained the four sounds described above in the order (buzz, tone, hiss, /i/)<sub>n</sub>. The duration of each sound was increased by 50% for successive groups of 30 Ss as follows (durations given for each sound, time for entire sequence four times the value given): 200, 300, 450, 670, 1,000 msec. In addition, a second group was tested with the 670-msec items using separate Ss and a different E. All Ss read the following instructions: "You will hear a train of four sounds presented in a certain sequence and repeated over and over. The sounds are: a buzz; a hiss; a tone; and the vowel 'ee' (as in the word 'beet'). When you have decided on the order, call out immediately and tell me what it is."

Starting to name the sequence with whichever sound he chose, S's chance of guessing the order of the other three sounds was one-sixth, so that the most probable number of correct responses was five for completely random guesses by the 30 Ss within each group. The results obtained are shown in Table 1. The numbers of correct responses were not significantly greater than chance until the duration of stimuli reached 670 msec. At this duration, 16 Ss identified the sequence correctly ( $p < .001$  by binomial expansion). A further increase of stimulus duration did not produce any appreciable increase in the number of correct responses. The attempt at replication using 670-msec stimulus durations was quite successful, with the same number of correct responses (16) as reported by the first group.

The distribution of response times within each group was skewed, with some Ss requiring over 200 sec, and for this reason, the data are presented as medians, with the distribution of response times indicated by the first and third quartiles (Q<sub>1</sub> and Q<sub>3</sub>). It can be seen that the longest response time was for the shortest duration (200 msec), although this rate provided the greatest number of

stimulus presentations in a given time. For item durations from 300 to 1,000 msec, the response time remained approximately constant, so that consistently fewer repetitions were required for response with increasing duration. Thus, at 300 msec, 22 repetitions were heard at the median response time, compared with 6 repetitions at 1,000 msec. Within groups, there appeared to be a tendency for correct responses to take longer. For stimulus durations of 450 msec and below, the number of correct identifications was at a level to be anticipated by chance, and as would be expected, there was no relation between accuracy and response time. Point biserial correlations showed that for the first of the two groups hearing stimuli lasting 670 msec, and for the 1,000-msec group, the listeners giving correct responses took significantly longer to respond ( $p < .05$ ). However, for the second group hearing items lasting 670 msec (the replication study), there was no significant correlation between response time and accuracy (response times were approximately equal for right and wrong answers). Care should be taken in interpreting the tendency for correct responses to take longer, for in the next study (Experiment 2), this tendency was reversed.

Table 1 indicates that the ability to identify temporal order with the procedure of this experiment is quite poor at all durations. It is surprising that accuracy no better than would be expected with random guessing was obtained at durations as long as 450 msec. It seemed possible that the method employed for obtaining responses could be interfering with accuracy. The next experiment was designed to eliminate some possible interfering factors and to allow a more efficient strategy for detecting temporal order.

*Experiment 2: Effect of duration of individual sounds upon card-ordering responses.* In the previous experiment, the vocal response consisted of names, most of which had only a slight acoustic resemblance to the sounds they represented, and hence the response could be subject to interference by the concurrent stimuli. Also, the procedure necessitated report of the entire sequence at one time. By requiring Ss to order cards bearing the typewritten names of the sounds, acoustic interference is avoided, and it is possible for S to deal effectively with successive partial responses. The method of arranging separate cards focuses attention on the relative positions of individual sounds: S can decide that one sound precedes or follows another, record this

decision with cards, and then focus his attention completely upon the next decision required to complete the description of the entire sequence.

The same stimulus tapes were used as in Experiment 1. The following instructions were read by Ss: "You will hear a train of four sounds presented in a certain sequence and repeated over and over. The sounds are described by the four separate cards lying on the table. You are to arrange the four cards in order of occurrence of the sounds. Starting with the card for one of the sounds, place it next to the number 1, and then arrange the remaining cards in order of occurrence of their sounds. When you have decided on the order you wish to report, call out and the experimenter will note the position of the cards."

Table 1 shows that for 200-msec durations, the number of correct responses (6) was not appreciably better than the most probable chance score of 5. However, for durations of 300 msec and above, all scores were significantly higher than chance at levels well above  $p < .001$  (binomial expansion). At 450-msec durations, the number of correct responses was highest: significantly higher than at 300 msec ( $p < .02$  by 2 by 2 contingency table), and higher than at 670 or 1,000 msec, although the superiority over these longer durations did not reach statistical significance.

Identification times were highest for 200 msec, and fell significantly with the increase in stimulus duration from 200 to 300 msec ( $p < .05$  by median test). As in Experiment 1, identification time remained approximately constant from 300 to 1,000 msec, and in this experiment, the median number of stimulus repetitions before responding varied from 25 to 6. Within each of the groups corresponding to stimulus durations from 300 to 1,000 msec, the response time was less for listeners responding accurately, although by point biserial correlation, this difference was significant only for the 450-msec group ( $p < .05$ ).

An analysis of wrong responses in both Experiments 1 and 2 revealed no favored incorrect order; answers seemed to be randomly distributed among the five possible incorrect responses.

It is obvious that sequence identification can be accomplished at much shorter stimulus durations for card-ordering than for oral responding with these four-item sequences. However, neither method permits identification at 200 msec, although perceptual ordering can be accomplished at considerably shorter durations for speech and music. The possibility exists that perceptual

processing of four items is too difficult with these unfamiliar sequences of sounds and that identification would be facilitated by changing to recycled sequences of three items. Further, one of the four nonrelated sounds was a portion of a steady-state vowel, and the presence of a speech-derived sound might invoke special mechanisms and contribute to confusion. In the next experiment, the recycled sequence consisted of the three nonspeech sounds employed in Experiments 1 and 2.

*Experiment 3: Sequence identification with three-item sequences.* The stimulus consisted of a 40-Hz square wave (buzz), a 1,000-Hz sinusoidal wave (tone), and a 2,000-Hz octave band noise (hiss), each item in the recycled sequence lasting 200 msec, with the order of items as given above. S could start with any of the three items, and since there are only two possibilities for arranging the remaining two sounds, he would be expected to guess correctly half the time. Because of the higher chance score than in Experiments 1 and 2, larger groups were used (50 Ss each) for each of the two conditions employed in this experiment.

One of the groups employed the oral response method with instructions similar to those used in Experiment 1, and the other group employed the card-ordering response method with instructions similar to those used in Experiment 2.

Performance for each group was not significantly different from what would be anticipated by chance: 23 out of 50 correct for oral response (median response time, 12.5 sec) and 27 out of 50 correct for card-ordering response (median response time, 35 sec).

The three-item sequence represents the smallest number of recycled items that can be used for detection of temporal order. Failure to achieve identification at 200 msec with both response procedures indicated a fundamental perceptual limitation for unpracticed Ss at these durations.

In the first three experiments, we have dealt with sequences consisting solely of nonrelated sounds. In the next experiment, two tones having frequencies corresponding to a musical interval were employed in a four-item sequence, together with the hiss and buzz used in Experiments 1, 2, and 3.

*Experiment 4: Identification of four-item sequences containing three types of sounds.* All sequences in this experiment consisted of the same four sounds: a hiss (2,000-Hz octave band noise), a buzz (40-Hz square wave), a high tone (1,000-Hz sine wave), and a low tone (796-Hz sine wave). The two tones were chosen so that they would

fall within the same octave, with the ratio of frequencies corresponding to a musical interval (a major third). Each item in the recycled sequences lasted 200 msec. Two groups of 30 Ss heard sequences with temporally contiguous tones [(high tone, low tone, buzz, hiss)<sub>n</sub>; (low tone, high tone, buzz, hiss)<sub>n</sub>], and two groups heard tones separated from each other by unrelated sounds [(high tone, hiss, low tone, buzz)<sub>n</sub>; (high tone, buzz, low tone, hiss)<sub>n</sub>]. Card-ordering responses were employed for each group, and the instructions were the same as those used in Experiment 2.

Since temporally contiguous notes of shorter duration can be ordered easily in music, it was anticipated that the sequences with one tone following the other directly might be identified more readily than the sequences with each tone both preceded and followed by a nonrelated sound.

Table 2 shows that correct identification of order was either near or at the most probable chance score of 5 correct for noncontiguous tones: 7 correct for (high tone, hiss, low tone, buzz)<sub>n</sub> and 5 correct for (high tone, buzz, low tone, hiss)<sub>n</sub>. Both sequences with contiguous tones had a frequency of correct responses significantly higher than chance:  $p < .05$  for 10 correct with (low tone, high tone, buzz, hiss)<sub>n</sub> and  $p < .001$  for 21 correct with (high tone, low tone, buzz, hiss)<sub>n</sub> (probabilities calculated from binomial expansion).

The median response times were similar except for (high tone, buzz, low tone, hiss)<sub>n</sub>, which was shorter than the other times. However, the difference was significant only when compared with the sequence (high tone, hiss, low tone, buzz)<sub>n</sub> ( $p < .05$  by median test).

An examination of the individual wrong responses for the noncontiguous tone sequences [(high tone, hiss, low tone, buzz)<sub>n</sub> and (high tone, buzz, low tone, hiss)<sub>n</sub>] was undertaken to determine if a perceptual bias toward temporal grouping of tones existed. The number of wrong answers in which the tones were reported to be contiguous was no greater than would be expected by chance. There seemed to be no favored incorrect responses.

An examination of the individual scores for the sequence of contiguous tones having the higher pitch first (21 correct responses) revealed that 6 of the 9 wrong responses involved solely transposition of the buzz and the hiss.

Identification of order with temporally contiguous tones was much more difficult (10 correct responses) when the lower pitch preceded the higher. Examination of individual responses showed that most of the 20

**Table 2**  
**Identification Through Card-Ordering of Repeated Four-Item**  
**Sequences\* Containing Two Related Sounds**

Sequence	Duration of Sounds (Msec)	Number Correct Out of 30	Response Time (Sec)	
			Median	Q1, Q3
(Hi, Hs, Lo, Bz) <sub>n</sub>	200	7	36	28, 54
(Hi, Bz, Lo, Hs) <sub>n</sub>	200	5	18.5	13, 36
(Lo, Hi, Bz, Hs) <sub>n</sub>	200	10	31.5	18, 61
(Hi, Lo, Bz, Hs) <sub>n</sub>	200	21	31	21, 46

\*The items were: Hiss (Hs)—2000-Hz octave band noise, Buzz (Bz)—40-Hz square wave, High Tone (Hi)—1000-Hz sine wave, Low Tone (Lo)—796-Hz sine wave.

wrong responses could be considered as simple transposition either within the tone pair [incorrect response (high tone, low tone, buzz, hiss)<sub>n</sub>—6 Ss], or within the nontone pair [incorrect response (low tone, high tone, hiss, buzz)<sub>n</sub>—8 Ss], with responses from only 6 Ss divided among the other 3 possible incorrect orders.

#### SUMMARY OF EXPERIMENTS 1-4

The three-item repeated sequences consisting of a hiss, tone, and buzz may be considered as the core or parent sequence from which all others employed in this experiment are derived. The temporal order of this three-item sequence could not be identified with either oral or card-ordering responses when the duration of each sound was 200 msec. The addition of another nonrelated sound (a steady state portion of the vowel /i/) produced a four-item sequence which was used to determine the thresholds for perception of order. Oral responses permitted identification significantly above chance at item durations of 670 msec and above, while card-ordering responses permitted identification at 300 msec and above.

When a second tone of different frequency was added to the core sounds (tone, hiss, buzz), a four-item sequence (each item lasting 200 msec) was created containing a pair of qualitatively similar sounds. Since different frequencies or pitches serve to distinguish the notes of melodies, it was anticipated that the correct ordering of tones might be accomplished more readily than the ordering of nonrelated sounds. This was found to be the case when the tones were temporally contiguous (i.e., the higher pitch tone, 1,000 Hz, either directly followed or directly preceded the lower pitch tone, 796 Hz). Identification for contiguous tone pairs was considerably better with falling pitch (high-low) than with rising pitch (low-high). When the tones were separated by nonrelated sounds [(high tone, hiss, low tone, buzz)<sub>n</sub> or (high tone, buzz, low tone, hiss)<sub>n</sub>], then identification of order was not better than for sequences containing

three of four nonrelated sounds having the same 200-msec duration.

We wish to emphasize that only unpracticed Ss were used in these studies and that each S made only a single judgment involving one of the sequences. The temporal confusion exhibited by some of our experimental groups does not seem to result from the inability of the untrained Ss to understand instructions or to identify the sounds with their names, for when the items were of sufficient duration, up to 80% of the Ss identified the sequences properly (see Table 1). While the results may be compared directly with the performance of other groups of unpracticed Ss, it is to be expected that different thresholds for temporal resolution would be found for the same stimuli with groups that have been trained or familiarized, either deliberately or incidentally, in the course of making multiple judgments. As indicated earlier, our preliminary observations indicate that some improvement occurs with little training. It is of interest that the values for practiced Ss reported by Thomas and his coworkers (Thomas, Hill, Carroll, & Garcia, 1970; Thomas, Cetti, & Chase, 1971) for repeated sequences derived from four steady state vowels were somewhat lower than those which had been reported by Warren (1968) for similar sequences of vowels presented to unpracticed Ss.

#### GENERAL DISCUSSION AND CONCLUSIONS

Our experiments on perception of temporal order in auditory sequences seem to weaken (or suggest modifications) of some widely held theories and generalizations. Also, these experiments link together diverse observations reported in the literature.

Firstly, we wish to question the accuracy of values frequently cited as representing the limits of resolution within sequences. Joos (1948), basing his conclusions mainly on speech, stated that there was a "time smear" due to limits imposed by processing mechanisms within the brain which produced temporal confusion and metathesis if successive sounds lasted between 50 and 100 msec. Similar

conclusions were reached by Stroud (1955), who employed concepts derived from information theory and suggested that there was a quantized "moment" for psychological time of about 100 msec, events within this quantum interval being effectively simultaneous and those above it being rank ordered in time. Recently (as mentioned earlier in this paper), numerous studies employing pairs of nonrelated sounds (such as hisses and tones) found temporal resolution possible at durations well below 100 msec. We have pointed out that this resolution for pairs of sounds seems to depend upon the ability to identify the sounds present at onset and termination of a series rather than perception of sequence per se.

Our experiments have indicated that identification of sequence for nonrelated sounds is not possible for untrained listeners hearing successive items each lasting 200 msec. Let us consider our sequence consisting of three recycled items. The individual sounds can easily be recognized and reported. To identify the order, the listener need only choose one of the sounds (say, the tone) and determine whether it was preceded by the hiss or the buzz. Since each sound is temporally contiguous with the other two, memory or storage time is minimal—indeed, it approaches zero. It is tempting to employ the post hoc explanation that recycling interferes with the identification of temporal order, but it should be recalled that recycled digits and isolated vowels can be ordered (Warren, 1968), as can the sequences described in this paper containing contiguous tones. More importantly, there exists evidence scattered through the literature indicating that accurate perception of order for extended sequences is limited to those consisting of phonemes or musical tones and, further, that within these unique perceptual continua of speech and music, rigid rules must be followed to permit identification of order. A brief sketch of some of these special rules follows.

Accurate ordering of a recycled series of individual vowels becomes more difficult when the sounds are deprived of selected characteristics (pauses, onset, and termination cues) associated with normal speech production (Warren, 1968). Perception of relative temporal position within a sentence is limited to those sounds forming the sentence: extraneous sounds, whether noises or other speech sounds, cannot be located accurately (Ladefoged, 1959; Garrett, Bever, & Fodor, 1966; Warren & Obusek, 1971).

An observation which may be of fundamental importance for

understanding the mechanisms for temporal ordering of items in speech has come from studies of identification times (Savin & Bever, 1970; Warren, 1971). It was found that identification of syllables and words occurs earlier than identification of their constituent phoneme clusters and individual phonemes. It seems clear that a simple Markovian model requiring initial identification of individual speech sounds followed by their arrangement into syllables or words is not applicable to perception of speech. It might be considered that the identification of syllables and phonemes reflects parallel or concurrent processes, with the less familiar activity of phonemic identification taking a little longer. However, Warren (1971) observed that semantic factors, which were manipulated to delay or accelerate the time required for identification of a syllable, added or subtracted equivalent intervals to the time required for identification of constituent phonemes. This evidence suggests to us that not only the perception of temporal order of phonemes in speech, but the very recognition of the presence of these sounds is an analytical process dependent upon prior identification of larger perceptual groupings.

In music and tonal sequences, characteristic rules for sequence perception have been observed. If the notes of one melody are alternated with another, and the two melodies are separated sufficiently in pitch, then the listener does not perceive the temporal contiguity of successive notes. Instead, the notes in each register are organized as separate sequences, and he "hears" two melodies (Ortmann, 1926). Incidentally, this principle has been used by J. S. Bach to make one instrument sound like two, each playing a separate melody. A related phenomenon has been reported by Miller and Heise (1950) and by Heise and Miller (1951), who found that if a tone in an extended sequence deviates in frequency from the neighboring tones by more than a "trill threshold" (equal to roughly one-seventh of an octave within the musical frequency range), it no longer seems to be part of the series, but "pops out." Kinney (1961) and Norman (1967) reported that tones differing from adjacent frequencies by more than the trill threshold could not be localized accurately within the tonal series. The method of recycled sequences

employed in the present paper has been used recently to measure the effect of frequency separation upon identification of order in tonal sequences. Using recycled six-item sequences consisting of 100-msec tones (350, 430, 550, 1,600, 2,000, 2,500 Hz), Bregman and Campbell (1971) found that the tones within the same octave were ordered accurately relative to each other, but that they could not be ordered relative to the tones in the remote octave. Thomas and Fitzgibbons (1971) employed sequences of four recycled 125-msec tones (frequencies not specified) and found that tones were ordered best when they were within a musical fourth.

Successive sounds of speech and music are linked by complex sets of rules characteristic of the linguistic and musical communities of which the listener is a member. Within these auditory continua, it is suggested that the order of successive items is deduced from the identifiable configurations which they form (i.e., identification of patterns precedes identification of the component items and their order). As we have seen, sequences of unrelated hisses, tones, and buzzes which do not form recognizable patterns, and also intrusive extraneous sounds occurring during speech or music, are not subject to temporal localization relative to other items. Further, if two conversations are in progress simultaneously, or two melodies are played in different registers, then only one of these competing sequences can be attended to at a time, and it is not possible to judge the relative temporal position of elements from the separate coherent sequences.

The inability to detect temporal order outside of the coherent patterns of speech (and music) may not correspond to a perceptual inadequacy, but may rather reflect mechanisms which, by excluding extraneous sounds, normally enhance the accuracy of pattern recognition within speech (and music).

#### REFERENCES

- AARONSON, D. Temporal factors in perception and short-term memory. *Psychological Bulletin*, 1967, 67, 130-144.
- BREGMAN, A. S., & CAMPBELL, J. Primary auditory stream segregation and perception of order in rapid sequences of tones. *Journal of Experimental Psychology*, 1971, 89, 244-249.
- EFRON, R. Temporal perception, aphasia and déjà vu. *Brain*, 1963, 86, 403-424.
- FAY, W. H. *Temporal sequence in the perception of speech*. The Hague: Mouton, 1966. Pp. 29ff.
- FRAISSE, P. *The psychology of time*. New York: Harper & Row, 1963. P. 89.
- GARRETT, M., BEVER, T., & FODOR, J. The active use of grammar in speech perception. *Perception & Psychophysics*, 1966, 1, 30-32.
- HEISE, G. A., & MILLER, G. A. An experimental study of auditory patterns. *American Journal of Psychology*, 1951, 64, 68-77.
- JOOS, M. Acoustic phonetics. *Language*, 1948, 24, 1-136 (Suppl.), (Language Monograph No. 23), pp. 75-76.
- KINNEY, J. A. S. Discrimination in auditory and visual patterns. *American Journal of Psychology*, 1961, 74, 529-541.
- LADEFOGED, P. The perception of speech. In NPL Symposium No. 10. *Mechanisation of thought processes*. Vol. 1. London: Her Majesty's Stationery Office, 1959. Pp. 398, 418.
- MALONE, R. L. Temporal ordering and speech identification abilities. *Journal of Speech & Hearing Research*, 1967, 10, 542-548.
- MILLER, G. A., & HEISE, G. A. The trill threshold. *Journal of the Acoustical Society of America*, 1950, 22, 637-638.
- NORMAN, D. A. Temporal confusions and limited capacity processors. *Acta Psychologica*, 1967, 27, 293-297.
- ORTMANN, O. On the melodic relativity of tones. *Psychological Monographs*, 1926, 35(1, Whole No. 162), p. 23.
- SAVIN, H. B., & BEVER, T. G. The nonperceptual reality of the phoneme. *Journal of Verbal Learning & Verbal Behavior*, 1970, 9, 295-302.
- STROUD, J. M. The fine structure of psychological time. In H. Quastler (Ed.), *Information theory in psychology*. Glencoe, Ill: Free Press, 1955.
- THOMAS, I. B., CETTI, R. P., & CHASE, P. W. Effect of silent intervals on the perception of temporal order for vowels. *Journal of the Acoustical Society of America*, 1971, 49, 84. (Abstract)
- THOMAS, I. B., & FITZGIBBONS, P. J. Temporal order and perceptual classes. *Journal of the Acoustical Society of America*, 1971, 50, 86-87. (Abstract)
- THOMAS, I. B., HILL, P. B., CARROLL, F. S., & GARCIA, B. Temporal order in the perception of vowels. *Journal of the Acoustical Society of America*, 1970, 48, 1010-1013.
- WARREN, R. M. Relation of verbal transformations to other perceptual phenomena. In Conference Publication No. 42, IEE/NPL Conference on Pattern Recognition. Teddington, England: Institution of Electrical Engineers, 1968 (Suppl., 8 pp.).
- WARREN, R. M. Identification times for phonemic components of graded complexity and for spelling of speech. *Perception & Psychophysics*, 1971, 9, 345-349.
- WARREN, R. M., & OBUSEK, C. J. Speech perception and phonemic restorations. *Perception & Psychophysics*, 1971, 9, 358-362.
- WARREN, R. M., OBUSEK, C. J., FARMER, R. M., & WARREN, R. P. Auditory sequence: Confusion of patterns other than speech or music. *Science*, 1969, 164, 586-587.
- WARREN, R. M., & WARREN, R. P. Auditory illusions and confusions. *Scientific American*, 1970, 223, 30-36.
- WINCKEL, F. *Music, sound and sensation*. New York: Dover, 1967. P. 54.

(Accepted for publication March 21, 1972.)