

## Auditory pattern recognition by untrained listeners\*

RICHARD M. WARREN

*University of Wisconsin-Milwaukee, Milwaukee, Wisconsin 53201*

Earlier work with unpracticed Ss has indicated that identification (naming) of the temporal order of components within repeated sequences consisting of three or four unrelated sounds cannot be accomplished when the item durations are 200 msec or less. In the present experiment, separate groups of 30 unpracticed Ss were required to tell whether alternated sequences, each consisting of reiterated presentations of the same three or four successive items, were in identical or permuted order. Naming of the order within the sequences was not required. Accuracy of same/different judgments was significantly better than chance when all items lasted 200 msec. Changing the duration of each item in one of the two sequences above or below 200 msec made the task more difficult. These results, together with other evidence, suggest that: (1) identification of order and recognition of auditory temporal patterns may represent fundamentally different processes, and (2) recognition may involve matching of "temporal templates."

A series of reports from this laboratory have indicated that the initial and terminal sounds of sequences are identified with especial ease, and that when such end effects are minimized, using recycled sequences consisting of three or four unrelated sounds (such as hisses, tones, and buzzes), the order cannot be identified by unpracticed listeners, even when the individual items last as long as 200 msec (Warren, 1968; Warren, Obusek, Farmer, & Warren, 1969; Warren, 1972; Warren & Obusek, 1972). This duration is considerably longer than the values of about 20-100 msec that had been accepted as the thresholds for perception of temporal order with various stimuli (see Fay, 1966; Warren & Obusek, 1972), and has suggested a new look at auditory temporal resolution.

Much of what appears to be direct naming of order may consist of initial recognition of the overall pattern followed by report of a learned descriptive label. This suggestion is consistent with the observation of Ladefoged and Broadbent (1960) that the overall "quality" of brief sequences needs to be familiar to listeners before a description of the order of the items is possible. There is recent evidence that in speech as well, pattern recognition precedes constituent analysis, so that identification of a syllable precedes identification of its component sounds (Savin & Bever, 1970; Warren, 1971). Observations with the phonemic restoration effect also indicate that phonemes are inferred, and not perceived directly (Warren & Obusek, 1971). If, indeed, much of what passes for perception of temporal order requires prior pattern recognition, perhaps formation of recognizable groupings is a fundamental process which

must precede temporal resolution of brief items. Then the difficulty in naming the order of sequences of hisses, tones, and buzzes would involve the inability to decompose the total perceptual pattern into successive components, even though the overall pattern itself could be recognizable and permuted orders distinguishable.

Somewhat similar conclusions concerning holistic auditory pattern perception were reached by Garner and his associates (Garner & Gottwald, 1967, 1968; Royer & Garner, 1970; Preusser, 1972) following experiments with rather different sequences consisting of extended patterns of dichotomous events (high tone, low tone). It was observed that organizational strategies of Ss changed with item durations, that a recognition task gave different results than an identification task, and that some sequences were perceived initially as a Gestalt, without direct identification of component items.

The present study measures recognition of unfamiliar auditory patterns at durations too brief to permit identification of order. Pairs of recycled sequences were used for same/different judgments to determine the characteristics and limits for discrimination of permuted orders of sounds.

### METHOD

Untrained Ss listened through headphones to two recycled sequences presented one at a time. Each contained the same items, and Ss were required to report whether these sequences were the same or different from each other. They did not have to identify the order, or even the component sounds. Each S made two judgments, one with a "same" pair and one with a "different" pair, half receiving the "same" first and half the "different" first. They could listen diotically to each member of the sequence pair for as long as they wished, and could switch from one to the other at will. There were no pauses introduced in switching sequences.

The primary reason for eliminating silence between paired sequences was the earlier finding that naming of order is facilitated greatly by even short pauses (Warren, 1972), suggesting that silence also might improve recognition of identity

\*This study was supported by grants from the National Science Foundation (GB-26459 and GB-36986X), and from the University of Wisconsin-Milwaukee Graduate School. Leong F. Tai provided valuable assistance with portions of the experimental work. A paper dealing with portions of this study was delivered at the 84th Meeting of the Acoustical Society of America, 1972.

or difference. Another reason for avoiding a silent interval is the possibility that the strength of the memory trace needed for comparing sequences could fade with time, making recognition dependent upon the particular selection of pause length separating the sequences to be compared.

It was found in preliminary experiments that a fixed relation between items in the two sequences at the moment of switching could influence judgments. In these experiments, the sequences for same/different judgments were recorded successively on the two channels of a stereo recorder. When all items in the two sequences had the same duration, a fixed relation between the channels was established inadvertently, so that any recorded sound on Channel 1 remained synchronous with a particular sound on Channel 2. This invariant relation between items in the two sequences heard upon switching appeared to influence the accuracy of same/different judgments in a manner dependent upon the fortuitous temporal alignment. It was possible to avoid this artifact by using a recording for only one sequence of the pair, generating the other sequence directly. In Experiment 1, whenever S was listening to the recorded channel, the recycling of the sequence controlled by the programming equipment halted, to be resumed when it was switched back on by S. This pause in sequencing on one channel eliminated any fixed relation between sounds on the separate channels. In Experiments 2 and 3, items in one sequence of the pair had one duration, and the items in the second sequence had another. The ratios of durations were not small integers, so that both sequences of the pairs could be recycled continuously without consistent cross-channel alignment of items.

#### Stimuli

The tones (1,000 and 796 Hz) and buzz (40-Hz square wave) were produced by electronic signal generators, and the noise band shaped by routing the output of a Gaussian noise generator through a filter adjusted to pass an octave band centered at 2,000 Hz. The vowel "ee" was prepared by cutting a 2-sec segment from a tape-recording of a longer steady production of the vowel held within 1 dB of a central value, splicing the segment into a loop and using the loop to rerecord a 20-min statement on a conventional reel of tape.

Ss heard pairs of sequences, each sequence consisting of three or four successive sounds repeated over and over in the same order without any pauses between items. The sequences forming a pair always contained the same sounds, but the temporal arrangement within the two sequences could be the same or different. Grason-Stadler Series 1200 programming equipment was used to generate all sequences. Transition from one item to the next in a sequence, and changes from one sequence to the other in a pair, were made using electronic switches set at rise/decay times of 5 msec. In Experiment 1, the comparison stimuli (Sequence B) were prerecorded on an Ampex AG 500 recorder, and the standard stimuli (Sequence A) were generated on-line during the experiment. Item duration in both sequences was 200 msec. Ss switched directly from one sequence to the other using a toggle switch. When switching from Sequence A, the programming timer stopped, and recommenced at the next item when switching back to A. The recording serving as Sequence B was free-running whether S was listening or not. In Experiments 2 and 3, items in Sequence A were 200 msec, items in Sequence B were at some fixed duration other than 200 msec, and both sequences were recycled continuously. In all three experiments, the starting positions within each sequence varied randomly for Ss within each experimental group.

#### Subjects

A total of 360 Ss (12 separate groups of 30 Ss) were recruited from the introductory psychology sections. Each received either \$1 or credit towards his course grade for participation, depending upon his wishes and the individual instructor's rules. Each S made only two judgments (one with "same" and one with "different" sequence pairs).

#### General Procedure

Ss were tested individually in a large double-walled audiometric room (IAC Model 1204A), in which both S and the E were seated at a table. All Ss read instructions informing them that they would be presented with a pair of sequences, each consisting of the same brief sounds which would be either in the same or in a different order. They were told that they could switch back and forth from one sequence to the other as often as they wished, and that when they decided whether the items in the sequence were in the same or in different orders, they were to inform the E. Any questions by Ss concerning the task were answered by rephrasing the appropriate portion of the instructions.

S then listened to a pair of sequences through matched TDH-49 headphones at a level of 80 dB re  $20 \mu\text{N/m}^2$ . The same items were presented in both sequences, and for half the Ss the order of items for first judgments was same, and for half, different. After S had indicated whether he thought the pair was same or different, he was presented with the other sequence pair, so that each S judged one same and one different pair. Following his second judgment, S was dismissed and not used further in sequence experiments.

## SPECIFIC PROCEDURES AND RESULTS

### Experiment 1: Identical Sequence Durations

The three separate groups of Ss used in this experiment each heard a different pair of sequences. At least one member of each pair had been used in earlier experiments dealing with the naming of the order of components (Warren & Warren, 1970; Warren & Obusek, 1972). Although it had been reported that identification of order with these sequences was not possible for untrained Ss at item durations of 200 msec, the results of the present experiment, summarized in Table 1, show that recognition (as measured by same/different judgments) can be made at these durations at levels significantly above chance. To illustrate the use of Table 1, let us consider Group 1 presented with recycled sequences containing three items: high tone, buzz, and noise. Each S judged whether a pair of sequences, consisting of HBN (high tone, buzz, noise, high tone, buzz, noise, etc.) for both Sequence A and Sequence B, had items in the same or in different order. Items in both sequences lasted 200 msec. The number of Ss correctly answering "same" was 19 (11 of the 30 Ss incorrectly answered "different"). Each of the Ss also received a pair consisting of HBN for Sequence A and HNB for Sequence B, and 27 of the Ss correctly described these sequences as "different." In order to minimize the possible effects of response bias (a tendency to choose one response rather than the other when uncertain), the same/different judgments were combined to yield 46 correct responses out of 60. The most probable chance score is 50% correct, and the value obtained is significantly above chance at  $p < .01$  by binomial expansion.

The sequences for Group 2 were similar to those in Group 1, except that a fourth item (796-Hz tone) was used in addition to the three items employed for the first group. It can be seen in Table 1 that there were a

**Table 1**  
Same/Different Judgments by Unpracticed Subjects (Five Separate Groups of 30 Subjects)

Group Number	Match	Repeated Sequence		Msec/Item		Number Correct	
		A	B	A	B	N = 30	Total (S + D)
Experiment I							
1	Same	HBN	HBN	200	200	19	
	Diff	HBN	HNB	200	200	27	46†
2	Same	HNLB	HNLB	200	200	17	
	Diff	HNLB	HBLN	200	200	22	39*
3	Same	HNEB	HNEB	200	200	20	
	Diff	HNEB	HBEN	200	200	24	44†
Experiment II							
4	Same	HNLB	HNLB	200	215	14	
	Diff	HNLB	HBLN	200	215	20	34
5	Same	HBN	HBN	200	670	16	
	Diff	HBN	HNB	200	670	14	30

Note—Stimuli were 1000-Hz high tone (H), 796-Hz low tone (L), 40-Hz square wave or buzz (B), 2000-Hz octave band noise (N), steady vowel "ee" (E).

\*Significant at  $p < .05$ , †significant at  $p < .01$ .

total of 39 correct responses, which was significantly better than chance ( $p < .05$ ).

Group 3 was similar to Group 2, except that the vowel "ee" replaced the 796-Hz tone, so that each of the four items was qualitatively different. It can be seen that the correct responses totaled 44 out of 60, which was significantly greater than chance ( $p < .01$ ).

#### Experiment 2: Different Sequence Durations

The items and orders used in Group 4 were the same as those used in Group 2 from Experiment 1, but the durations of items in Sequence B were all changed to 215 msec. Since fixed alignment of items across sequences was not possible with these different item durations, the special programming for changing alignment used in Experiment 1 was unnecessary, and both Sequences A and B were prerecorded. It was found that the total number of correct responses dropped to 34, which is not significantly better than chance, and not significantly different from Group 2 in Experiment 1.

Group 5 had the same three items presented in the same orders as in Group 1. While items in Sequence A remained at 200 msec, items in Sequence B were lengthened to 670 msec [at this longer duration, it has been found that unpracticed Ss can name the order of components for these items (Warren & Warren, 1970)]. Table 1 shows that listeners found it impossible to make same/different judgments with this degree of temporal mismatch of the recorded sequences—the total number of correct responses is the same as the most probable chance score.

#### Experiment 3: Temporal Mismatch Limits for Recognition

The items and orders used for all seven groups in this experiment were identical with those employed for Group 3 in Experiment 1 [Sequence A: tone, noise,

"ee," buzz; Sequence B: tone, noise, "ee," buzz (same) or tone, buzz, "ee," noise (different)]. Each group had the duration of items in Sequence A at 200 msec (as in Experiment 1), but the durations of items in Sequence B had a different value for each of the groups: 127, 160, 215, 315, 415, 515, and 600 msec, respectively. Sequences A and B were recycled continuously, as in Experiment 2, and both sequences were generated on-line. The results are summarized in Table 2 and Fig. 1. It can be seen that the total number of correct responses was significantly better than chance only for item durations in Sequence B of 215 and 315 msec (as well as the duration of 200 msec used in Experiment 1). Decreases in duration below 200 msec and increases in duration above 215 msec produced monotonic decreases in accuracy of recognition of temporal order, as shown in Fig. 1.

#### DISCUSSION

When the accuracy of the same and of the different judgments were compared within each of the 12 groups used in the three experiments, significant differences were found only for Group 1 in Experiment 1 (different more accurate than same at  $p < .05$  using 2 by 2 contingency table). However, it is quite possible that relative accuracy of recognition of "same" and "different" orders is confounded with biases towards making one type of judgment when uncertain for some sequences. Such biases would not be expected to increase the total number of correct responses, and so the combined same/different scores were taken as a measure of the ability to recognize temporal order under the conditions employed.

Experiment 1 demonstrates that recognition of identity or difference in temporal order is possible at durations reported to be too short to permit

Table 2  
Results of Experiment III: Effect of Duration Mismatch on Same/Different Judgments With Recycled Sequences Consisting of the Same Recycled Four-Item Sequences (Separate Group of 30 Untrained Subjects for Each Duration)

	Duration of Items in Sequence B							
	127	160	[200]*	215	315	415	515	600
Same Correct	16	20	[20]	24	18	17	16	15
Different Correct	17	18	[24]	21	21	20	18	18
Total Correct	33	38	[44]†	45†	39**	37	34	33

Note—Sequence A always had items lasting 200 msec (order: tone, noise, "ee," buzz). Sequence B had durations as shown above with order identical to Sequence A for "same" and with position of noise and buzz interchanged for "different." Each of the Ss made two same/different judgments (one judgment involving identical and one judgment involving nonidentical orders) for a group total of 60 judgments.

\*Group 3 from Experiment 1; \*\*significant at  $p < .05$ ; †significant at  $p < .01$ .

identification (naming) of order of the components within these sequences by untrained Ss (Warren & Warren, 1970; Warren & Obusek, 1972). Does this difference in performance for the two tasks indicate a fundamental difference between mechanisms for recognizing individual permuted orders, on the one hand, and identifying (naming) of order, on the other?

With sequences containing several items, it might be argued that the identification of the order of components and the recognition of the pattern have a common basis: that partial identification of order involving a portion of the total sequence could be responsible for recognition of the pattern, so that correct same/different judgments would be possible without the ability to identify completely the entire sequence. Accepting for the moment this common basis, since Ss could make accurate same/different judgments with recycled items in Group 1 of Experiment 1, they should be able to make partial identification of order. But with a three-item recycled sequence, any partial identification of order involving the relative order of two sounds is equivalent to complete specification for all three. For example, if S concentrated on the noise and the buzz, and identified the noise as being first, S has by

so doing identified the position of the tone (it must follow the buzz and precede the noise), and the entire sequence has been characterized. Yet, in experiments dealing with identification of order with the same three-item sequences (Warren & Warren, 1970; Warren & Obusek, 1972), Ss did not perform at levels above chance. The procedure for naming order in these experiments (arranging cards containing the names of the items in the proper order) was designed to keep information storage minimal (Ss needed to arrange only two of the three cards in proper order while listening to the sequence). Any comparison of orders for same/different judgments would require additional storage, and should be a more difficult task, suggesting that the relative ease of making same/different judgments is the consequence of mechanisms which do not require such ordering.

Other evidence that same/different judgments do not involve partial identification of order has been provided by an unpublished experiment by Byrnes and Warren. Untrained Ss tried to match each of the six possible orders of four recycled tones (each of the tones had a different frequency, but the same duration and intensity), using a comparison array consisting of all six

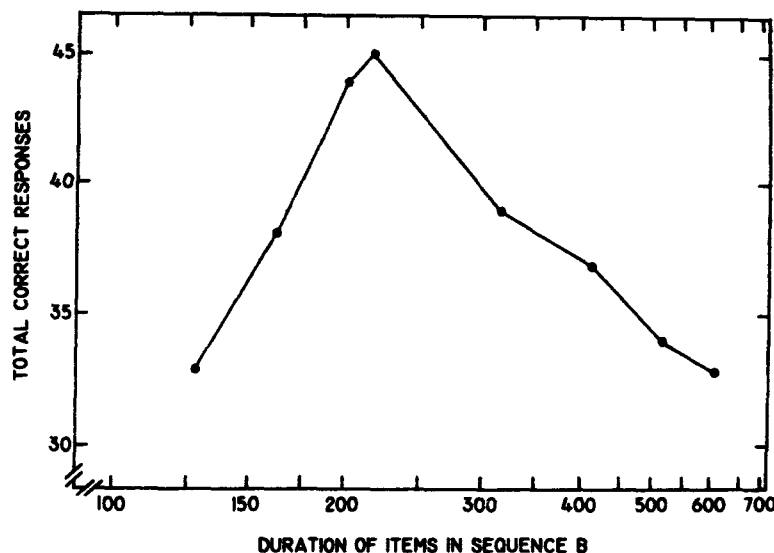


Fig. 1. Shape of an auditory temporal template. The total numbers of correct same and different responses (maximum score is 60 correct) for pairs of recycled four-item sequences are shown as a function of the logarithm of item duration in milliseconds of Sequence B. Items composing Sequence A were always 200 msec. Scores for each duration are from separate groups of 30 Ss.

permuted orders of the tones. The same Ss also tried, at a separate session, to identify the order by arranging four cards describing the sounds along a continuum from lowest to highest pitch. Performance for recognition (matching) was better than that corresponding to a partial identification of order. Also, it was found that the effects of changing either frequency separation of components or item durations upon accuracy of responses were quite different for recognition of patterns and for identification of order, indicating a basic difference in the processes underlying these tasks.

How, then, are permuted orders of sequences recognized if the orders of individual components cannot be identified? It may be that some aggregate perceptual quality characterizes the experimental sequences and permits recognition, much as a recognizable melody can characterize a sequence of notes. If the overall pattern or form is used for recognition, an interesting problem arises which will become clearer if we turn for a moment to vision. One can recognize a face despite changes in the visual angle subtended. It is not necessary that the pattern of stimulation evoked by a retinal image match a stored template as a casting matches its mold. A smaller portrait may be placed alongside an enlargement of the same negative, and recognition of identity or difference is easy despite size disparity. The temporal dimension in hearing is often considered as analogous to the spatial dimension in vision. On this basis, we would expect that if same or different judgments could be made accurately with duration of items in recycled sequences of 200 msec, then expanding the duration of items in one of the sequences should not decrease the accuracy of recognition. On the other hand, it might be that recognition of order in hearing requires the matching of a temporal template.

Experiments 2 and 3 tested a "temporal template" model for auditory pattern recognition. The sounds were those used in Experiment 1, but with changed duration of items in one of the two sequences of each pair. In Experiment 2, Group 5 had the duration of Sequence B for the three-item patterns used in Group 1 expanded to 670 msec, with Sequence A remaining at 200 msec. At the longer duration, the temporal order can be named readily (Warren & Obusek, 1972). If comparisons of component orders were responsible for same/different recognition (rather than global recognition or temporal template matching), accuracy should not decline when durations in one sequence were expanded to permit accurate identification of order. However, the combined scores of the Ss presented with this temporal mismatch were at the most probable chance value. The difference between scores in Group 1 and Group 5 was significant at  $p < .01$  (2 by 2 contingency table).

Group 4 in Experiment 2 had a slight mismatch of the four-item sequences used in Group 2 in Experiment 1. When all items were 200 msec in Group 2, combined same/different judgments totaled 39 correct, which was

just significantly better than chance at the .05 level. Lengthening the duration of items in Sequence B to 215 msec produced a combined score of 34, which was not significantly different from chance or from the value of 39.

While Group 5 in Experiment 2 indicated that a gross temporal difference of over 300% prevented accurate comparison of temporal order, it did not establish the limits for matching of temporal templates. Experiment 3 systematically explored the accuracy of same/different judgments with various temporal mismatches. Sequence A was always 200 msec, and Sequence B had values from 127 through 600 msec. Figure 1 shows the manner in which temporal mismatch influences the accuracy of same/different judgments. The abscissa corresponds to the logarithm of durations of items in Sequence B, resulting in steeper slopes for longer durations than would be produced with linear representation of time. Even so, it can be seen that the slope of the curve for item durations longer than 200 msec is less steep than that for the shorter durations. Taking the percentage change from 200 msec necessary to reduce accuracy of identification half way to the most probable chance value (i.e., midway between 44 and 30, or 37 correct), we find a marked asymmetry: a reduction in duration of 25% produces the same effect as an increase of 100%. It appears that the extent of temporal mismatch was more critical for the shorter than for the longer durations.

Turning to familiar patterns of sounds, of course recognition is not restricted to a perfectly rigid set of characteristic durations. Some degree of flexibility is tolerated in speech and music. But the temporal variability allowed in auditory pattern recognition may be quite small when compared with the spatial variability allowing object recognition at various distances in vision.

## CONCLUSIONS

It appears that recognition of sequences as measured by same/different judgments does not necessarily involve partial identification of the temporal order of components, but can be based upon a type of holistic perception. The results of this study (together with other evidence based upon verbal and tonal sequences) indicate that recognition of overall patterns of sounds can occur at item durations too brief to allow direct naming of order of components. The stored "temporal templates" employed for recognition of sequences have characteristic limits for the extent of temporal mismatch permitting pattern recognition.

## REFERENCES

- Fay, W. H. *Temporal sequence in the perception of speech*. The Hague: Mouton, 1966.  
 Garner, W. R., & Gottwald, R. L. Some perceptual factors in the learning of sequential patterns of binary events. *Journal of Verbal Learning & Verbal Behavior*, 1967, 6, 582-589.

- Garner, W. R., & Gottwald, R. L. The perception and learning of temporal patterns. *Quarterly Journal of Experimental Psychology*, 1968, 20, 97-109.
- Ladefoged, P., & Broadbent, D. E. Perception of sequence in auditory events. *Quarterly Journal of Experimental Psychology*, 1960, 12, 162-170.
- Preusser, D. The effect of structure and rate on the recognition and description of auditory temporal patterns. *Perception & Psychophysics*, 1972, 11, 233-240.
- Royer, F. L., & Garner, W. R. Perceptual organization of nine-element auditory temporal patterns. *Perception & Psychophysics*, 1970, 7, 115-120.
- Savin, H. B., & Bever, T. G. The nonperceptual reality of the phoneme. *Journal of Verbal Learning & Verbal Behavior*, 1970, 9, 295-302.
- Warren, R. M. Relation of the verbal transformation effect to other perceptual phenomena. In Conference Publication No. 42, IEE/NPL Conference on Pattern Recognition, Institution of Electrical Engineers, Teddington, England, 1968 (Supplement, 8 pp.).
- Warren, R. M. Identification time for phonemic components of graded complexity and for spelling of speech. *Perception & Psychophysics*, 1971, 9, 358-363.
- Warren, R. M. Perception of temporal order: Special rules for the initial and terminal sounds of sequences. *Journal of the Acoustical Society of America*, 1972, 52, 167. (Abstract)
- Warren, R. M., & Obusek, C. J. Speech perception and phonemic restorations. *Perception & Psychophysics*, 1971, 9, 358-363.
- Warren, R. M., & Obusek, C. J. Identification of temporal order within auditory sequences. *Perception & Psychophysics*, 1972, 12, 86-90.
- 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 (December), 223, 30-36.

(Received for publication May 25, 1973;  
revision accepted January 2, 1974.)