

RESEARCH ON SPOKEN LANGUAGE PROCESSING
Progress Report No. 19 (1993-1994)
Indiana University

**Speaker Variability and Attention in an
Auditory Serial Reaction Time Task¹**

Carl W. Turner, Scott E. Lively, & David B. Pisoni

*Speech Research Laboratory
Department of Psychology
Indiana University
Bloomington, Indiana 47405*

¹This research was supported by NIH Research Grant DC-00111 and NIDCD Training Grant DC-00012 to Indiana University in Bloomington, IN.

Abstract

Two experiments were conducted to determine the effects of speaker variability, attention, and level of pattern redundancy on implicit learning in an auditory serial reaction time task. Subjects in both experiments responded to auditorily presented words given four visual alternatives. In Experiment 1, the auditory words were spoken by a 16 speaker ensemble, in a randomly determined order. In Experiment 2, the speakers were grouped by 12 trial sequence. There were no effects of pattern or speaker presentation on serial reaction times, but there was an effect of explicit knowledge of the repeating pattern. The reasons for the lack of effects on reaction time in these auditory SRT experiments are discussed.

Speaker Variability and Attention in an Auditory Serial Reaction Time Task

The accurate perception of spoken words often depends on the speaker characteristics in the speaker ensemble, or the set of speakers to which the listener is exposed. Early research in speech perception showed that listeners were slower and less accurate in identifying words spoken by a series of talkers compared to words spoken by a single talker (Creelman, 1957; Peters, 1955; Summerfield & Haggard, 1973). These early studies did not, however, appear to lead to an extended program of work on speaker variability.

Recently, a series of studies on speaker variability, speech intelligibility, and memory have been conducted that extend these early experiments and account for them in terms of current theories of speech processing. In one study, listeners ability to identify spoken words was affected by the number of speakers in the speaker ensemble (Mullennix, Pisoni, & Martin, 1989). For words presented in the clear and in noise, performance was better for words that were produced by a single talker than words produced by multiple talkers; trial-to-trial variability in the speakers voice affected recognition performance. This was explained as the listener's need to "recalibrate" their auditory perceptual mechanism each time a new voice was encountered.

Voice information may in fact demand attentional resources, even when the speaker's voice is irrelevant to the task at hand. In a study on the role of attention and speaker characteristics, subjects were asked to classify words either by initial phoneme or by speaker (Mullennix & Pisoni, 1990). Interference as measured by reaction time to correct classification increased as the speaker- and word-set increased, suggesting that the stimulus in the irrelevant dimension affected performance in the relevant dimension.

Word perception is of course a resource demanding process. Taken together, these recent findings suggest that voice characteristics may also demand attentional resources, even when a speaker's voice is not apparently part of the task. Listeners may incidentally encode voice information in the course of speech perception, a process that may take resources from the primary task of understanding spoken words. If so, then the presentation of words by different speakers could affect performance on secondary tasks that also require attention.

The Serial Reaction Time Task

A series of recent studies in visual perception and serial pattern learning have employed a novel technique with which to measure learning and attention and provides the instrument of study for the present experiments. Nissen and Bullemer (1987) presented subjects with a target in one of four locations on a computer screen. Subjects pressed computer keys based on the location of the target. Unknown to the subjects in the experimental conditions, the location of the target was varied based on a complex 10-sequence pattern. Reaction times decreased throughout the test trials for subjects who viewed patterned targets; there was no decrease in reaction times in a random-pattern control condition. In a second experiment, subjects who performed a secondary tone counting task did not improve on the primary task. Nissen and Bullemer interpreted these results to mean that the secondary task required attentional resources, and that full attention is necessary for serial pattern learning to occur.

Cohen, Ivry, and Keele (1990) showed that serial pattern learning can occur under conditions of divided attention provided that the serial pattern has unique and repeating sequences. If there is redundancy in the pattern then learning can occur in the presence of a secondary task. Thus, the sequence

CADBADCBA (which may represent positions of targets on a screen in four given locations) the position A is always followed by D. In this way Cohen et al. attempted to demonstrate the importance of both attention and the underlying pattern to be learned.

Currently there is some disagreement over the causes for failure of subjects to learn the serial patterns when attention is diverted by a secondary task. Other studies have claimed that interference in perceptual grouping of the patterns, rather than division of attentional capacity, accounts for failure to learn serial patterns (Frensch & Miner, in press; Stadler, 1993). One finding is that serial learning can occur in the presence of distractor tasks, but only in cases where there is redundancy in the serial pattern. Rather than attention, the most important variable seems to be consistent groupings within the serial patterns (Stadler, in press). However, there remains controversy over the precise nature of attention and implicit pattern learning: "In sum, attentional orienting programs may play an important role in learning serial reaction time and similar tasks, although the experimental evidence so far is inconclusive. It is also unclear to what degree other implicit learning tasks might depend on these attentional mechanisms." (Seger, in press, pp. 176-177)

The present experiments adapted the visual serial pattern learning design to an acoustic environment. Briefly, subjects saw four words on a computer screen and simultaneously heard one of the words. The task was to press one of four buttons corresponding to the spoken word. There were three levels of statistical structure to the patterns: random, low redundancy, and high redundancy. If serial learning does progress independent of an attentional manipulation, then the voice manipulations should have no effect on learning, and learning should occur based strictly on the level of the statistical structure of the pattern (Stadler, 1992).

There is the question of speaker variability and attention. If changing the speaker from trial to trial demands more attention than hearing a single voice then we expect to find differences between groups should be based on that variable. If there is linkage between speaker characteristics and pattern learning then it should be apparent in the interactions between voice and the level of redundancy of the pattern. If, on the other hand, implicit learning of serial patterns occurs independent of attentional manipulations, then there should be no interaction between voice manipulations and the level of pattern.

Experiment 1

A persistent question about serial reaction time experiments deals with the extent to which subjects are consciously aware of the repeating pattern and can use their knowledge to facilitate performance on the task. Conscious knowledge of the repeating pattern used to order the stimuli can result in unusually short reaction times, thereby affecting the interpretation of the results (Shanks & St. John, in press). Conscious knowledge of the serial pattern may be difficult to assess but a number of tests have been used, including open ended questionnaire and a generate task (Nissen & Bullemer, 1987). In this task subjects are required to generate the sequence in which the stimuli will appear. Rather than responding to the presentation of a stimulus, subjects are asked to predict the location of the next stimulus on a trial-by-trial basis. Feedback is given after each trial. This test reinstates conditions in the serial reaction time test and is considered to be a more sensitive measure of awareness than a retrospective questionnaire (Nissen & Bullemer, 1987; Shanks & St. John, in press).

In one auditory version of the serial reaction time task, subjects pressed keys in response to the spoken words "one, two, three, four" (Palmeri, Lively, & Pisoni, 1991). Subjects heard words spoken in the same voice or by different voices to the words in a 10-sequence repeating pattern. There was no effect

of speaker variability. As measured by retrospective questionnaires, 116 out of 176 subjects in the experimental conditions reported either "some" or "full" knowledge of the serial pattern at the end of the experiment. Furthermore, decreases in reaction times across four blocks of trials closely followed the levels of explicit knowledge of the task. Analysis of the generate task showed that subjects reached asymptotic explicit knowledge of the serial pattern on the first block of generate trials. It was thought that this task was extremely simple for the subjects to perform and allowed them to attend to other aspects of the task, including the repeating pattern of button presses.

In the present Experiment 1, subjects heard one of 288 possible words in any one of 16 voices on any given trial. The four alternatives presented on the computer screen differed by only one phoneme, making discrimination of the target word somewhat difficult. It was thought that this design would minimize the likelihood that subjects would not notice the repeating pattern. Explicit knowledge of the serial pattern was assessed by a questionnaire and by the generate task.

Method

Subjects. Forty subjects participated in this experiment. Thirteen subjects served in a random pattern condition, 14 in a low redundancy pattern condition, and 13 in a high redundancy pattern condition. All were recruited from the student population at Indiana University or from the surrounding community and paid \$5.00 for their participation. Subjects were tested in groups of 6 or fewer. All were native speakers of English who reported no speech or hearing disorders.

Stimuli. All spoken words were single syllable words from the Modified Rhyme Test (MRT). Each word was recorded on audiotape and digitized by a 12-bit analog-to-digital converter. The RMS amplitude levels for all words were digitally equated. All words were spoken by 16 different speakers (8 male and 8 female).

The 288 words selected from the MRT list consisted of 48 word groups of six words each. In each word group, the six words varied from each other on the first phoneme of the word or on the last phoneme. Twenty-four word groups varied on the first phoneme, 24 on the last. On a given trial, subjects viewed four words in a word group printed in capital letters at the bottom of a computer screen, and simultaneously heard one of the words presented through headphones.

Design. Subjects were randomly assigned to one of three groups based on serial pattern: random, low redundancy, or high redundancy. The random sequences were entirely random, with the constraint that the same response not occur twice in a row and that the same number of responses on each button (24) occur within the 96 trial block. The low redundancy pattern was ABADBCDACBDC, in which the letter A corresponded to button 1 on a response box, B to button 2, and so on. This pattern has the lowest level of statistical structure possible in a 12-trial sequence (Stadler, 1993). The high redundancy pattern was ABADBC DACDBC. Note that in the high redundancy pattern the series of events DBC occurs twice within the sequence, as does the pair CD. Each 12 trial sequence was administered eight times in each 96 trial block.

Subjects performed four blocks of 96 test trials, followed by a block of generate trials. In each block, two words were selected from each of the 48 word groups as the target word in a trial for 96 total auditory stimuli. On every trial, three alternatives were selected randomly from the remaining five words for presentation on the screen. A different speaker was selected for each word such that every speaker in the 16-speaker ensemble spoke six words within a block. Within a block of trials the order of the speakers

was randomized. Finally, the words were arranged for display on the computer screen so that the target word appeared in a random location (for the random pattern condition) or in a predetermined location (for the low and high redundancy conditions). All of the 288 words from the MRT list were used at least once during the experiment, and many were used twice.

Procedure. Subjects were tested in sound-treated booths equipped with a computer screen, headphones, and a four-button response box. Buttons on the response box were labeled "1" to "4" from left to right. In the serial reaction time test, four words were presented on the screen and a word presented auditorily. Subjects were instructed to respond as quickly as possible to the auditory word by pressing one of the four buttons corresponding to the words on the screen. After each trial, a light above the correct button was lit to indicate the correct response. The deadline for responding was two seconds, after which the screen went blank and the next trial began. The time between trials was 1500 msec.

After each trial block subjects were given a two-minute rest period. Upon completing the four blocks of the serial reaction time test, a written questionnaire was administered to assess subjects' awareness of the repeating pattern. The questions were "1. Did you notice anything about the task? 2. Did you ever notice a pattern or repeating sequence? (YES or NO)" and "3. If there was a pattern, could you reproduce it? (Use the numbers 1,2,3,4 to represent the button numbers)." After the questionnaires were collected, subjects began the generate test. In this test, subjects were presented with four words on the screen, and asked to press a button corresponding to the word they thought they would hear. After all subjects had responded, the auditory word was presented and the correct response light was turned on. Subjects were given a deadline of six seconds in which to respond, after which the next trial began. As in the previous four test blocks, there were 96 trials in the generate block.

Results

Subjects in the two pattern conditions were first scored on their questionnaires with the aim of dividing them into groups according to full knowledge (11 or 12 responses correct), partial knowledge (5 to 10 correct) or no knowledge (less than 5 correct). However, none of the subjects in either group showed knowledge of the pattern; most claimed not to have noticed a pattern of button responses at all, although several commented that the same words were spoken more than once, and that the alternatives rhymed with the target word.

Subjects made few errors in responding during the serial reaction test, as shown in Table 1. Errors may have occurred due to the difficulty of discriminating some of the words, or from a speed/accuracy tradeoff. No further analysis was conducted on the numbers of correct judgments in the serial reaction test.

Insert Table 1 about here

The median reaction time for correct trials within each 12-trial sequence was computed for each subject (Willingham, Nissen, & Bullemer, 1989). These data were averaged across the eight sequences within each block, and a 3 x 4 (Pattern x Trial Block) ANOVA was conducted. There was no effect of Pattern on reaction times, $F(2, 37) = 2.1, p = .13$, and no Pattern x Trial Block interaction. There was a main effect of Trial Block on reaction times, $F(3, 111) = 15.9, MSe = 1539.6, p < .0001$. As seen in Figure 1, subjects were faster in the second through the fourth trial blocks than in the first.

Table 1

Percentage correct responses in the serial reaction time test in Experiment 1

Pattern	<i>n</i>	Trial Block				Total
		1	2	3	4	
Random	13	94.1	94.8	93.9	94.9	94.3
Low Redundancy	14	93.1	96.4	96.6	95.2	95.3
High Redundancy	13	93.8	93.5	93.2	94.4	93.7

Insert Figure 1 about here

An ANOVA was performed on the percentage correct for the three groups in Trial Block 5, the generate task. As expected, there was a main effect of Pattern, $F(2, 37) = 9.1$, $MSe = 100.2$, $p < .001$. The random group scored only 27.7% correct on the generate task, which was at a chance level of performance. Planned comparisons showed that both the high redundancy ($M = 37.8\%$) and low redundancy ($M = 44.8\%$) groups performed better on the generate task than the random group, $t(25) = 4.5$, $p < .0001$ and $t(24) = 2.7$, $p < .05$, respectively. The two patterned groups did not reliably differ from each other, $t(25) = 1.5$, $p > .10$.

Discussion

Although subjects in all three groups showed decreases in reaction times across the four blocks, the two patterned groups did not differ from the random pattern group. Thus, the small decrease in reaction times between the first and second trial blocks must be attributed to task learning unrelated to the underlying pattern of button responses.

The task was sufficiently attention-demanding that none of the subjects in the pattern groups showed even partial knowledge of the pattern as indicated by questionnaire scores. There was evidence of explicit knowledge as measured by the generate task, but it is unclear whether this knowledge was extant at the beginning of the generate block or whether it was acquired in the course of performing the generate task itself.

Subjects' reactions to the question about patterns or repeating sequences were varied. Of those who responded positively, several remarked on the similarity of the spoken words to the alternatives. Also, most subjects claimed to have heard only four or five voices during the experiment, suggesting that discriminating between the voices was somewhat difficult.

Experiment 2

In Experiment 1, subjects were presented with words spoken by different speakers in a 16-speaker ensemble, with the speaker varying randomly from trial to trial. This style of presentation forced listeners to devote attentional resources in an effort to recalibrate to each speaker's voice and may have been responsible for the subjects' inability to learn the underlying patterns. It was hypothesized that presenting words in the same voice on successive trials would reduce attentional demands and improve performance on the serial reaction time task.

Experiment 2 used the same stimuli but the voices were arranged such that the same voice spoke each word in a 12 trial sequence. In this arrangement, there were two possible sources of facilitation. Presenting words by the same speaker might reduce the attentional demands of the task by reducing the listener's need to recalibrate to different voices on every trial. On the other hand, tying a single voice to a 12 trial pattern might have the effect of making the underlying pattern more salient to the listeners by delimiting the beginning and end of the pattern. This would reflect the importance of subjects' organization of the material for serial reaction time learning.

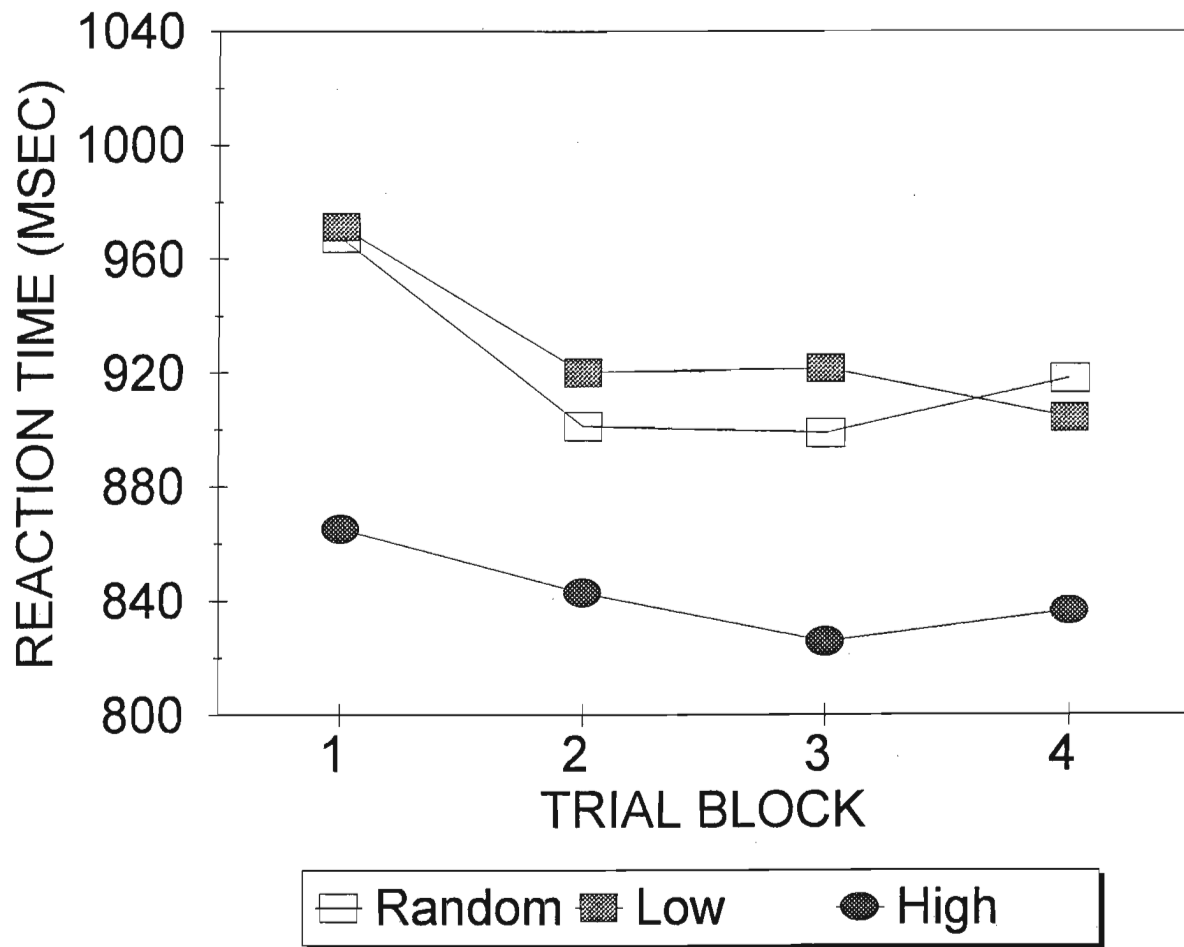


Figure 1. Mean of the median reaction times per sequence for correct responses in the serial reaction time task in Experiment 1.

If grouping speakers by sequence in the serial reaction task facilitates performance by reducing attentional demands, then there should be an overall effect on reaction times both for the experimental and control groups. On the other hand, if facilitation occurs due to better organization of the pattern, then improved performance should be found only in the experimental groups, not the control groups for which there is no underlying pattern.

Method

Thirty-three subjects participated in this experiment; 13 in a random pattern condition, 10 in a low redundancy condition and 10 in a high redundancy condition. All were recruited from the student population at Indiana University or the surrounding community and paid \$5.00 for their participation. All were native speakers of English who reported no speech or hearing disorders.

The stimuli, design, and procedure were identical to the previous experiment with one exception. As before, each block of 96 trials consisted of eight 12 trial sequences. In the present experiment, the words spoken on each trial within a 12 trial sequence was presented in a single voice. In the four blocks of the serial reaction time task every voice in the 16 voice ensemble was used twice. The words in the generate task were also organized in this manner; the voices used in the eight sequences were randomly selected from the 16 voice ensemble.

Upon completion of the serial reaction task, subjects were given the written questionnaire, then performed the generate task.

Results

Subjects in the two pattern groups were scored on their questionnaires and divided into full, partial, and no knowledge groups. Two subjects in the low redundancy group and one subject in the high redundancy group showed complete knowledge of the pattern. That these subjects were able to employ this knowledge during the serial reaction test was shown in their mean correct reaction times in Trial Block 4: less than 300 msec, far faster than the group means at the same point in the experiment. These subjects were not considered in further analysis of reaction times or percentage correct in the generate task. Three other subjects, two in the low redundancy group and one in the high redundancy group, showed partial knowledge of the pattern. Their reaction times in the serial reaction test were consistent with the other subjects in their respective groups, and were therefore included in the analysis.

As in Experiment 1, subjects made few errors in responding during the serial reaction test. See Table 2. No further analysis was conducted on the numbers of correct judgments in the serial reaction test.

Insert Table 2 about here

The median reaction time for correct trials within each sequence was computed for each subject. These data were averaged across the eight sequences within each block, and a 3 x 4 ANOVA was conducted. There was again no effect of Pattern on reaction times, $F(2, 27) = 1.9, p = .16$. There was a main effect of Trial Block on reaction times, $F(3, 81) = 5.56, MSe = 2,083.9, p < .01$. There was a Pattern x Trial Block interaction, $F(6, 81) = 2.7, p < .05$. As shown in Figure 2, subjects in the random and high

Table 2

Percentage correct responses in the serial reaction time test in Experiment 2

Pattern	<i>n</i>	Trial Block				Total
		1	2	3	4	
Random	13	95.4	97.0	95.0	94.7	95.5
Low Redundancy	10	95.0	95.7	93.4	95.3	94.9
High Redundancy	10	93.7	94.4	94.2	95.3	94.4

redundancy groups were increasingly faster to respond from the first to the fourth trial blocks; subjects in the low redundancy group, however, did not follow this trend.

Insert Figure 2 about here

An ANOVA was run on the percentage correct for the three groups in the generate task. There was a main effect of Pattern, $F(2, 27) = 11.7$, $MSe = 210.8$, $p < .0001$. The random group scored 27.6% compared to 54.9% for the high redundancy group and 54.4% for the low redundancy group. Planned comparisons showed that both the high and low redundancy groups scored higher in the generate task than the random group, $t(20) = 5.5$, $p < .0001$ and $t(19) = 4.2$, $p < .0001$, respectively.

In Experiment 1, a question arose over the validity of the generate task in measuring subjects' explicit level of knowledge of the pattern during the serial reaction task. It has been suggested that the generate task is a more sensitive measure of explicit knowledge than the written questionnaire (Shanks & St. John, in press). The relatively high level of correct trials in the generate task in Experiment 2 coupled with the low level of improvement in the serial reaction task indicate that the generate task may have overestimated subjects' explicit knowledge. In order to look more closely at the data from the generate task, subjects' mean percentage of correct trials were broken out into the eight 12-trial sequences (see Figure 3). A regression analysis was performed for each group.

Insert Figure 3 about here

The results clearly showed that subjects in the two pattern conditions improved on the generate task. Simply averaging across all eight sequences in the trial block yielded an inflated estimate of their pattern knowledge. The regression equation for the high redundancy group was $y = 37 + 3.97x$, $r^2 = .63$. The equation for the low redundancy group was $y = 45.9 + 1.88x$, $r^2 = .33$. The equation for the random group was $y = 30.2 + (-0.57)x$, $r^2 = .35$.

Discussion

As seen in Experiment 1, subjects in showed decreases in reaction times during the serial reaction task. Unlike Experiment 1, subjects in the high redundancy and random groups seemed not to have reached an asymptotic level of performance, and the low redundancy group was actually slower in the fourth trial block than in the first. These data obtain for a very small sample, however; due to the process of dropping subjects with full knowledge of the patterns, there were only eight subjects in the low redundancy condition and nine in the high redundancy condition.

The outcome of the generate task for assessing explicit knowledge was of interest. Although there has been some concern that the retrospective questionnaire is not sufficiently sensitive for detecting explicit knowledge, the generate task of the kind used in these experiments may in fact produce an inflated estimate of conscious knowledge. Other recognition-type tasks for assessing awareness of serial patterns may give a more accurate estimate of conscious knowledge (Stadler, in press).

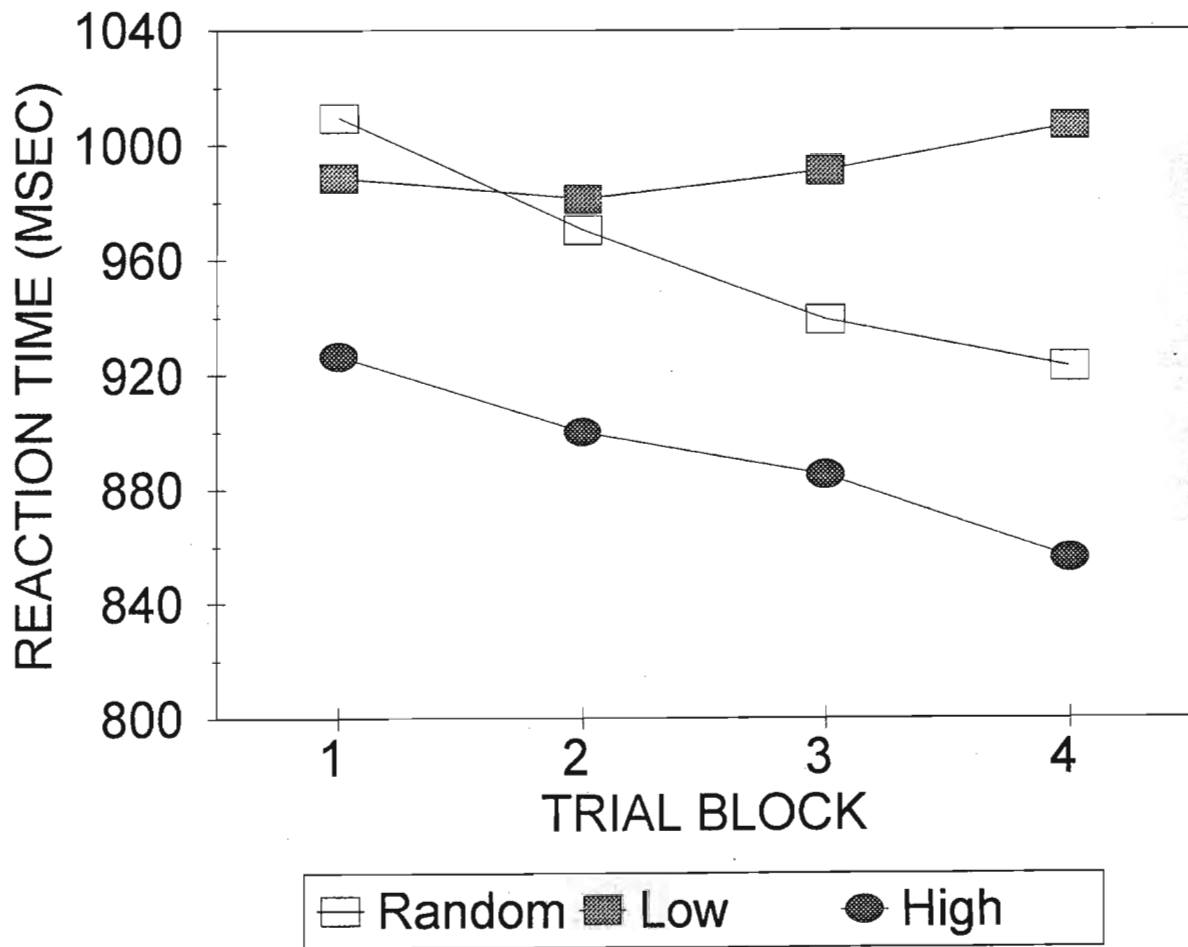


Figure 2. Mean of the median reaction times per sequence for correct responses in the serial reaction time task in Experiment 2.

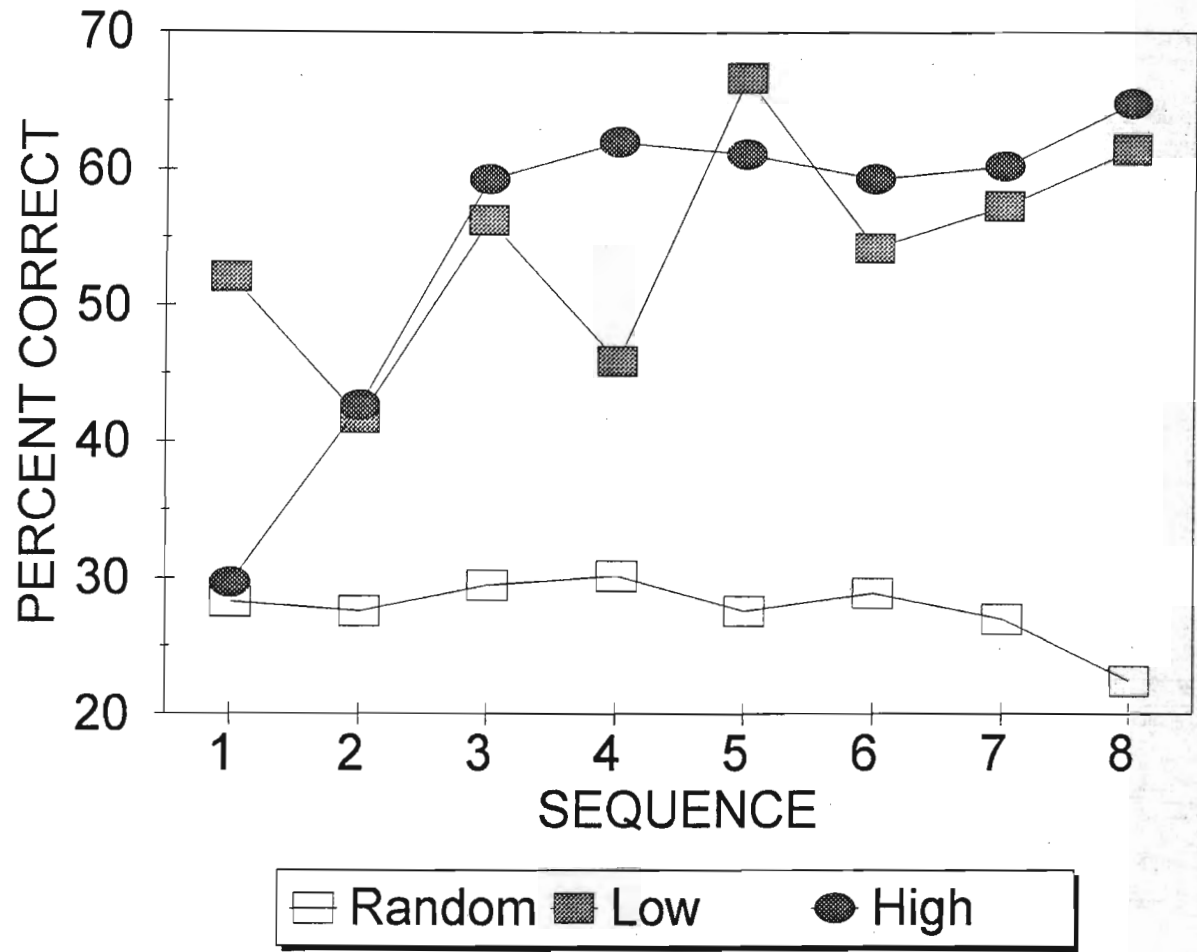


Figure 3. Mean percentage correct responses in the generate task in Experiment 2.

General Discussion

The failure of the pattern manipulation to provide evidence of serial pattern learning was somewhat surprising, given the success of the technique in earlier auditory SRT experiments (Hartman, Knopman, & Nissen, 1989; Palmeri et al, 1991). The present experiments, with the large number of stimuli, the relative difficulty of mapping a spoken word onto a response, and the longer pattern (12 vs. 10 trials) may all have had an effect on subjects' performance. This design also lacked a spatial component that is characteristic of visual SRT experiments; the targets did not appear in substantially different locations on the screen but as one of four alternatives on a line of text. Subjects reading the line in a typical left-to-right fashion would gain little from implicitly learning the location of the target words on the screen.

The results of the experiments did not confirm or disconfirm the importance of organization in serial pattern learning. There was no evidence of serial learning in either experiment, so no conclusions can be drawn. Other attempts to make serial patterns more salient, and therefore learnable, by delimiting the patterns (e.g., Lewicki, Hill, & Bizot, 1988) have been successful.

Grouping the speakers by sequence in Experiment 2 seemed to have an effect on performance. Two of the three groups showed increased facilitation through the course of the four trial blocks, compared to Experiment 1 where subjects reached asymptotic performance fairly quickly. Extending the design to five or more trial blocks may have provided more definitive data.

References

- Cohen, A., Ivry, R. I., & Keele, S. W. (1990). Attention and structure in sequence learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*, 17-30.
- Creelman, C. D. (1957). Case of the unknown talker. *Journal of the Acoustical Society of America*, *29*, 655.
- Frensch, P. A., & Miner, C. S. (in press). Effects of presentation rate and individual differences in short-term memory capacity on an indirect measure of serial learning. *Memory & Cognition*.
- Hartman, M., Knopman, D. S., & Nissen, M. J. (1989). Implicit learning of new verbal associations. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*, 1070-1082.
- Lewicki, P., Hill, T., & Bizot, H. (1988). Acquisition of procedural knowledge about a pattern of stimuli that cannot be articulated. *Cognitive Psychology*, *20*, 24-37.
- Mullennix, J. W., and Pisoni, D. B. (1990). Stimulus variability and processing dependencies in speech perception. *Perception & Psychophysics*, *47*, 379-390.
- Mullennix, J. W., Pisoni, D. B., & Martin, C. S. (1989). Some effects of talker variability on spoken word recognition. *Journal of the Acoustical Society of America*, *85*, 365-378.
- Nissen, M. J., & Bullemer, P. (1987). Attentional requirements of learning: Evidence from performance measures. *Cognitive Psychology*, *19*, 1-32.
- Palmeri, T. J., Lively, S. E., & Pisoni, D. B. (1991). Implicit learning of auditory sequences: A first report. *Research on Speech Perception: Progress Report No. 17* (pp. 379-395). Indiana University, Bloomington, IN.
- Peters, R. W. (1955). The relative intelligibility of single-voice and multiple-voice messages under various conditions of noise. *Joint Project Report No. 56* (pp. 1-9). U. S. Naval School of Aviation Medicine, Pensacola, FL.
- Seger, C. A. (in press). Implicit learning. *Psychological Bulletin*.
- Shanks, D. R., & St. John, M. F. (in press). Characteristics of dissociable human learning systems. *Behavioral and Brain Sciences*.
- Stadler, M. A. (1992). Statistical structure and implicit serial learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*, 318-327.
- Stadler, M. A. (1993). Implicit serial learning: Questions inspired by Hebb (1961). *Memory & Cognition*, *21*, 819-827.
- Stadler, M. A. (in press). The role of attention in implicit learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*.
- Summerfield, Q., & Haggard, M. P. (1973). Vocal tract normalization as demonstrated by reaction times. *Report Of Speech Research In Progress*, *2*(2) (pp. 12-23). Queens University of Belfast.
- Willingham, D. B., Nissen, M. J., & Bullemer, P. (1989). On the development of procedural knowledge. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*, 1047-1060.