

Automaticity and the detection of speech

JOHN W. MULLENNIX

Wayne State University, Detroit, Michigan

and

JAMES R. SAWUSCH and LAURIE F. GARRISON

State University of New York, Buffalo, New York

The development of automatic perceptual responses to speech stimuli was examined. In the first experiment, phoneme-monitoring performance for speech syllables was examined under conditions in which stimulus-to-response mapping and memory load were manipulated. The results indicated that automaticity develops under consistent-mapping conditions. In the second experiment, a dual-task procedure was combined with mapping and selective attention manipulations in order to examine the development of automaticity across single- and multiple-channel conditions. The results indicated that performance under consistently mapped training conditions was interfered with by dividing attention across multiple channels of input. It is concluded that there may be differences in the way that automaticity develops across visual and auditory modalities and that these differences need to be examined more closely.

A number of viewpoints about attention incorporate a distinction between automatic and controlled processing (e.g., LaBerge, 1975; Logan, 1978, 1979; Neumann, 1984; Posner & Snyder, 1975; Schneider, 1985; Shiffrin & Schneider, 1977). Automatic processes are described as unavoidable, occurring without capacity limitations, awareness, or intention, highly efficient, and resistant to modification (LaBerge, 1981, p. 173). In contrast, controlled processes are described as relatively slow, generally serial, and capacity-limited, and they require attention (Shiffrin & Schneider, 1977). Some theories place a great deal of importance on the manner in which automatic processing develops, with an emphasis on examining the consequences of training or practice on the development of automaticity. One theory falling into this category is the two-process theory of attention and human information processing proposed by Schneider and Shiffrin (1977; Shiffrin & Schneider, 1977). Schneider and Shiffrin showed that the manner in which the stimulus is mapped to the response determines whether a task can become automatized. They demonstrated that after a period of training under consistent-mapping (CM) conditions, increases in memory load and stimulus display size have little effect on performance, whereas under varied-mapping (VM) conditions, both factors adversely affected performance. On the basis of their results, Schneider and Shiffrin suggested that automatic, capacity-free detection processes develop as a function of consistent training.

In recent years, the appropriateness of this theory has been tested by extending the initial findings of Schneider and Shiffrin (1977) to different experimental paradigms. One method that has been used is to examine automaticity under dual-task conditions (Fisk & Schneider, 1983; Logan, 1978, 1979; Schneider & Detweiler, 1988; Schneider & Fisk, 1982a, 1982b; Strayer & Kramer, 1990). In these studies, situations were constructed where an "automatized" task was combined with a concurrent task that required capacity or resources. The logic behind this manipulation is that if automatic processing develops under CM conditions and, as a result, requires little or no processing capacity, then the addition of a concurrent capacity-limited task should not interfere with performance of the automatized task, and vice versa. Indeed, this pattern of results has been found for various visually oriented tasks (e.g., see Logan, 1979; Schneider & Fisk, 1982a). These results provide further evidence for automaticity under consistent-training conditions and indicate that the dual-task method is useful for assessing the attentional requirements of automatic and controlled processing.

In general, the research devoted to assessing the nature of automaticity and its development is based largely upon evidence obtained in visual search and detection tasks, category search tasks, and Sternberg (1966) memory tasks. As a consequence, many theories of attention and automaticity (LaBerge, 1975; Logan, 1978, 1979; Neumann, 1984; Posner & Snyder, 1975; Schneider, 1985; Shiffrin & Schneider, 1977) are based upon a restricted domain of research evidence obtained in tasks requiring the processing of visual stimuli. When considering theories of attention, it is important to extend the empirical findings concerning automaticity to the auditory modality. Auditory processing differs in some re-

The research reported here was supported by NINCDS Grant 5 R01 NS19653-02 to the State University of New York at Buffalo and by NIH Training Grant NS-07134-09 to Indiana University in Bloomington, IN. The authors would like to thank Mary Czerwinski and Howard Nusbaum for comments and suggestions on earlier versions of the manuscript. Reprint requests may be sent to the first author at the Department of Psychology, Wayne State University, 71 W. Warren St., Detroit, MI 48202.

spects from visual processing, being inherently temporal with information processed through two spatial channels (ears). Because of modality-specific processing differences, it cannot be assumed that the allocation of attention and the development of automaticity are identical in both visual and auditory modalities.

There is some research in the perception of speech related to this issue. Poltrock, Lansman, and Hunt (1982) adapted the experimental paradigm of Schneider and Shiffrin (1977) for use with spoken word stimuli. They used an auditory target-detection task, with subjects monitoring dichotic streams of spoken letter names (i.e., A, B, G, H, I, L, R, U). Subjects responded to the target letters that never appeared simultaneously on both channels. The results Poltrock et al. obtained resembled the findings of Schneider and Shiffrin in that performance was superior for consistently mapped stimuli and memory load had a differentially greater effect on performance under VM conditions. Poltrock et al. therefore concluded that automatic attentional responses developed to auditory letter-word stimuli under CM conditions.

With regard to the results of Poltrock et al. (1982), it must be noted that they presented the words dichotically with both items in the same talker voice. We know that the dichotic presentation of stimuli in the same voice can lead to perceptual fusions, blends, or other confusions (Cutting, 1976) that can affect perception in various ways. More important, however, is the fact that their subjects had to monitor both channels of input. This means that Poltrock et al.'s results reflect processing when attention is divided over two auditory spatial channels. This situation brings up an interesting question concerning automaticity in speech processing. That is, is there any processing "cost" to the development of automaticity when speech stimuli are monitored across two spatial channels of input instead of one channel of input? At first glance, Poltrock et al.'s results would seem to suggest that the development of automaticity was little affected by monitoring multiple channels of input, because processing under those conditions became automatized anyway. But, their results cannot directly address this question because they did not assess the development of automaticity in a single-channel situation for comparison. In addition, other studies in the literature examining speech processing under CM conditions across multiple channels of input present mixed results (Moray, 1975; Ostry, Moray, & Marks, 1976). Moray (1975) reported that, under CM training conditions, dichotic targets across two channels were processed equally as well as dichotic targets were when one channel was selectively attended to and as well as binaural (one-channel) targets were. However, Moray (1975) and Ostry et al. (1976) reported perceptual interference when two targets occurred simultaneously across channels. Moray (1975) attributed this interference to response-level factors, implying that both channels were perceived equally well. However, it is still unclear whether limitations in perceptual factors or response factors were actually responsible for their results.

Given these findings with speech stimuli, the answer to the question of processing cost, with regard to automaticity and multiple-channel processing, remains unclear. Since the processing of spatial channels of temporal information is quite important in audition, it would be informative to examine in more detail whether the allocation of attention to multiple-input channels interferes with the development of automaticity. The purpose of this study was to construct an experimental situation in which the likelihood that this could be observed would be maximized.

In our first experiment, we wished to extend Poltrock et al.'s (1982) findings by investigating whether automatic responses develop to speech syllables under single-channel (instead of dichotic multiple-channel) conditions. If the results are similar to Poltrock et al., then further evidence will have been obtained suggesting that the principles underlying the development of automaticity are similar for speech tasks and visual tasks.

In our second experiment, we investigated automaticity and multiple-channel processing in a more thorough manner than was previously attempted. We designed a task that combined training (CM vs. VM) and dual-task methodologies in a manner that allowed the comparison of single- and multiple-channel processing under various training/load conditions. If there is no processing cost for monitoring multiple input channels under CM conditions, then concurrent processing loads should have little effect on perception, and there should be no difference between single- and multiple-channel CM task conditions.

EXPERIMENT 1

The purpose of Experiment 1 was to examine whether or not automatic perceptual responses to speech syllables on a single channel develop as a result of CM training. The experimental paradigm of Schneider and Shiffrin (1977; Shiffrin & Schneider, 1977) was adapted for use and modified for investigating the development of automaticity for speech targets. In Experiment 1, we assessed reaction time (RT) monitoring performance for consonant-vowel (CV) speech syllables under conditions where S-R mapping and memory load were manipulated. If responses to speech syllables become automatized, two results should be observed. First, overall RT performance should be better under CM conditions than under VM conditions. Schneider and Shiffrin (1977; Shiffrin & Schneider, 1977) observed gradual, steady decreases in RTs over training sessions under CM conditions, whereas RTs asymptoted quickly and improved very little under VM conditions. The second result that should be observed concerns the effect of memory load on performance. Under CM conditions, the effect of memory load should become negligible and disappear over the course of training. However, under VM conditions, the effect of memory load should only be slightly reduced, with a substantial effect of memory load still present at the end of training. This pattern of results would parallel the findings obtained by Schneider and Shiffrin with visually presented stimuli and

by Poltrock et al. (1982) with auditory stimuli. On the other hand, if this pattern of results is not found, then the hypothesis that responses to speech syllables become automatized would be weakened.

Method

Subjects. The subjects were 8 paid volunteers obtained from the State University of New York at Buffalo and the surrounding university community. All subjects were right-handed, native speakers of English who reported no history of a speech or hearing disorder. Four subjects took part in 20 1-h sessions, and 4 subjects took part in 30 1-h sessions (one session per day). Each subject was paid \$4 per hour for participating.

Stimuli. The stimuli consisted of 24 naturally produced stop-consonant-vowel syllables obtained from a male talker. The CV syllables were selected by pairing the six stop consonants /b/, /d/, /g/, /p/, /t/, and /k/ with the vowels /a/, /ae/, /i/, and /u/. Each stimulus was produced within the carrier phrase, "Please say _____ for me," with the blank corresponding to the target CV syllable. The stimuli were lowpass filtered at 4.8 kHz and converted to digital form via a 12-bit analog-to-digital converter at a 10-kHz sampling rate. The CV syllables were digitally edited from the carrier phrase to produce the final stimulus materials used in the experiment. The vowel offsets of the stimuli were edited and the amplitude falloff was adjusted in order to produce a uniform length of 305 msec per stimulus.

Procedure. There were four experimental conditions: CM one-target, CM three-target, VM one-target, and VM three-target. Each subject received S-R mapping as a within-subject variable and memory set size as a between-subject variable. The subjects were equally divided and randomly assigned to two groups. Each subject in the one-target condition was run for 20 consecutive daily sessions, with the first 10 sessions comprising one mapping condition and the second 10 sessions comprising the other mapping condition. Each subject in the three-target condition was run for 30 consecutive sessions, with the first 15 sessions comprising one mapping condition and the second 15 sessions comprising the other mapping condition. The number of sessions was determined by monitoring performance on a day-to-day basis in order to determine how many sessions were required for performance to stabilize in both CM and VM conditions. Stable performance was defined as a non-significant decrease in RT occurring on any two consecutive sessions. As a result of using this criteria, 10 sessions were conducted under one-target conditions and 15 sessions were conducted under three-target conditions.¹

The experimental procedure involved a phoneme-monitoring task. The subjects were instructed to monitor lists of speech syllables presented over headphones and to make an appropriate response to any of the target consonants. The target items were randomly interspersed among a number of distractor stimuli within each stimulus list. For example, if the target consonant was "b," the target syllables [ba], [bae], [bi], and [bu] were randomly presented among the other CVs. The subjects were instructed to respond to the targets by pushing a button on a response box as quickly as possible whenever they heard a syllable containing a target consonant. The stimuli were presented binaurally over matched and calibrated TDH-39 headphones, with a stimulus presented every 1,005 msec. RTs were recorded from stimulus onset.

Memory set size was manipulated by requiring the subjects to respond to either one consonant or any of three consonants. For the one-target set, the subjects monitored the CV syllables and responded only to the syllables containing the syllable-initial target consonant. For the three-target set, the subjects monitored the CV syllables and responded whenever a syllable occurred that contained any of three different syllable-initial target consonants. To

prevent the subjects from making responses to all three targets on the basis of one phonetic feature only (e.g., place of articulation or voicing), the three-target sets were composed of consonants that did not have one feature common to all. For example, one three-target set that was used consisted of the consonants "b," "d," and "k," with all three consonants differing in place and only two consonants sharing a common voicing feature. Target sets sharing a distinctive phonetic feature, such as voicing (i.e., "b," "d," and "g") were not used.

The mapping variable was manipulated by presenting the targets in CM sets or VM sets. For the CM condition, the S-R mapping remained the same over all sessions. The subjects responded to the same consonant target during the one-target sessions or to the same three consonant targets during the three-target sessions throughout all of the lists presented over all of the CM sessions. The distractors consisted of all the other stimuli not containing target consonants, with the distractors also remaining the same throughout all CM sessions. Assignment of phonemes to CM memory set was randomized across subjects.

For the VM condition, the S-R mapping changed for each block of stimulus lists over all the sessions. In this condition, for the one-target sessions, each subject received a different consonant target at the beginning of each block of nine stimulus lists, and, for the three-target sessions, each subject received three different consonant targets at the beginning of each block. For example, in the one-target condition, one subject responded to "b" targets in the first block of stimulus lists, "k" targets for the second block of lists, and so on. For the three-target condition, a subject responded to "b," "d," and "k" targets in the first block, "p," "t," and "g" targets in the second block, and so on. The presentation order of consonant targets in the VM conditions was randomized across blocks of lists and across sessions.

Before the presentation of each list of stimuli, the target consonant set was displayed on a CRT screen in front of the subject and remained on screen during stimulus-list presentation. Each subject was presented with nine lists of 20 trials, for a total of 180 randomized trials in a block with six blocks per session, 36 targets per block, with a 5:1 ratio of distractors to targets. No more than two target syllables occurred in succession within a particular list, and no targets occurred in the first or last positions in the list. Each list contained anywhere from two to six targets, for an average of four targets per list. Each subject initiated the presentation of each list by pressing a button on the response box. Stimulus presentation and data collection were controlled on-line by a PDP-11/34A computer. RTs were recorded from stimulus onset.

Results and Discussion

For each subject, mean percent hits, mean percent false alarms, and mean RTs were calculated over sessions for each of the mapping and memory set conditions. RTs were analyzed for correct responses (hits) only.

Figure 1 displays mean RTs for the two groups of subjects over sessions as a function of memory set size and mapping condition. Inspection of the figure shows that the pattern of RT performance over sessions differs substantially between conditions. In the VM three-target condition, performance appears to stabilize immediately and improve very little over the course of training. For CM one-target, CM three-target, and VM one-target conditions, RT gradually decreased over sessions and, after the stabilization criterion was reached, was probably not at asymptote. We have plotted the 10 sessions for the one-target group with Sessions 6-15 for the three-target group

Experiment 1 (RT Over Sessions)

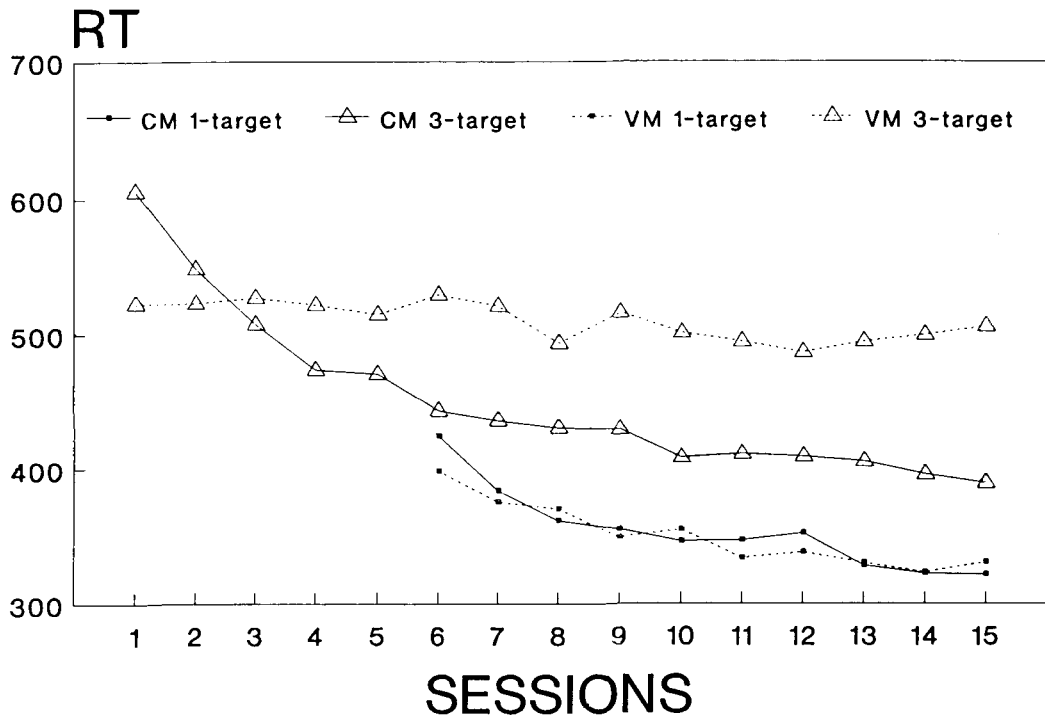


Figure 1. Mean reaction time (RT) plotted over sessions for consistent- and varied-mapping groups as a function of memory set size for Experiment 1. RT is plotted over 10 sessions for the one-target conditions and 15 sessions for the three-target conditions.

so that the last sessions, when subjects met our stabilization criterion, are aligned.

To assess performance at the end of training, an analysis of the data was conducted on the subjects' performance averaged over the last two sessions. An analysis of variance (ANOVA) was run, with the between-subject factor of memory set size (one vs. three) and the within-subject factor of mapping (CM vs. VM). Table 1 displays the data as a function of memory set size and mapping in terms of mean RTs, mean percent hits, and mean percent false alarms averaged across the last two sessions for each experimental condition.

A significant main effect of mapping on RT was obtained [$F(1,6) = 10.2$, $MS_e = 1,297.7$, $p < .02$]. RT was faster in the CM conditions than in the VM conditions. A significant main effect of memory set size on RT was also observed [$F(1,6) = 13.3$, $MS_e = 4,567.1$, $p < .01$]. RT was faster in the one-target conditions than in the three-target conditions. In addition, a significant interaction of mapping \times memory set size was obtained on RT [$F(1,6) = 8.5$, $MS_e = 1,297.97$, $p < .03$]. Post hoc tests of the interaction revealed that RT performance in the VM three-target condition differed significantly from all other three conditions, RT performance in the CM three-target condition differed from both one-target con-

ditions and the VM three-target condition, and RT performance between the one-target conditions did not differ. In other words, RT performance in the CM three-target condition was intermediate between the VM three-target condition and the two one-target conditions.

The difference in RT between one-target and three-target conditions was examined as a function of mapping. Under VM conditions, the RT advantage for the one-target condition over the three-target condition was 123.9 msec on Session 1 and 175.3 msec on the last session (Session 10 for one-target, Session 15 for three-target). Thus, the difference in RT between memory set size conditions actually increased over training under VM conditions. In contrast, under CM conditions, the RT advantage for the one-target condition over the three-target condition was 180.8 msec on Session 1 and 67.9 msec on the last session. In this case, the effect of memory load diminished substantially over training.

Separate two-way ANOVAs for the factors of mapping and memory set size were also conducted on hits and false alarms. A significant main effect of mapping was obtained for hits [$F(1,6) = 6.1$, $MS_e = 0.0003$, $p < .05$]. Fewer hits were observed in the VM condition than in the CM condition. No other significant main effects or interactions were observed for hits or false alarms. The results of these

Table 1
Mean Reaction Times (in Milliseconds), Hits (% Correct), and False Alarms (% FAs) for Consistent and Varied Mapping Groups for Memory Set Size of One and Three in Experiment 1

Group	RT	Hits	FAs
One-Target Condition			
CM	322.0	99.8	1.1
VM	327.1	99.4	1.0
Three-Target Condition			
CM	392.8	99.1	0.8
VM	502.9	95.3	1.7

Note—CM = consistent mapping; VM = varied mapping.

analyses suggest that the differences in RT performance found in Experiment 1 were not due to speed-accuracy tradeoffs in the data.

Overall, performance was better for CM conditions than for VM conditions, and the effects of memory load were much greater for VM conditions than for CM conditions at the end of training. These results fit previous criteria for the development of automaticity (see Schneider & Shiffrin, 1977). One observation concerning the data is that a significant residual difference in RT at the end of training between the one- and three-target CM conditions was observed. A likely reason for this difference is between-group variability, since memory set size was manipulated between subjects in this experiment. However, similar residuals between CM mapping conditions have been documented for both visually based tasks and auditory-based tasks (Poltrock et al., 1982; Schneider & Shiffrin, 1977).

A second observation is that the number of false alarms did not significantly differ across conditions. With regard to the possibility of negative transfer from CM to VM sets, one would expect that if negative transfer had occurred, then the false-alarm rate would substantially increase. However, false-alarm rates did not significantly change. This result is consistent with our assertion that negative transfer between CM and VM conditions did not affect the overall pattern of results.

Overall, the pattern of results for Experiment 1 indicates that automatic perceptual responses to speech syllables develop under CM training conditions. When making direct comparisons between our experiment and other studies, the manner in which our practice schedules were constructed and the way in which the mapping variable was manipulated should be taken into account (see Note 1).

EXPERIMENT 2

The results of Experiment 1 extend the findings of Poltrock et al. (1982) to monitoring a single channel of auditory input. Given that automaticity develops for a single channel of input, we now turn our focus to multiple-channel-input conditions. In Experiment 2, we created an experimental situation in which any processing cost to the development of automaticity due to multiple-channel monitoring could be examined under the rubric of CM/VM training conditions.

To accomplish this, the dual-task methodology previously used for investigating automaticity in visual tasks (Fisk & Schneider, 1983; Logan, 1978, 1979; Schneider & Detweiler, 1988; Schneider & Fisk, 1982a, 1982b; Strayer & Kramer, 1990) was adapted for use with speech. Since we wanted to compare single-channel monitoring with multiple-channel monitoring, we combined this task with an attentional manipulation (see Bookbinder & Osman, 1979; Nusbaum, 1981). For example, Nusbaum (1981; described in Nusbaum & Schwab, 1986) conducted a series of experiments examining the perceptual monitoring of isolated vowels and CV speech syllables under conditions of focused and divided attention. Using a dual-task procedure, Nusbaum (1981) demonstrated that monitoring performance deteriorated when attention was divided across two input channels (ears) relative to a one-channel control. These results indicate that the perception of speech across input channels was capacity-limited. We decided to adapt the attentional manipulation of Nusbaum in order to create conditions under which single- and multiple-channel monitoring could be compared under CM and VM training conditions.

The procedure consisted of a dual-task phoneme-monitoring paradigm, with listeners monitoring for and responding to appropriate speech (phoneme) targets. The primary task in all conditions was to monitor for CM target consonants. However, in one condition, the secondary task was to monitor for a CM target (different from the primary), whereas, in another condition, the secondary task was to monitor for a VM target. This dual-task situation allowed us to assess the effect of a concurrent processing load on the CM primary task. If automaticity develops under CM conditions, one would expect that neither a CM nor a VM concurrent processing load would have an effect on the CM primary task.

In addition, targets were monitored under conditions of selective attention to one channel of input or divided attention across two channels of input. When attention was focused on one channel, both primary and secondary targets occurred on one channel of input. When attention was divided, secondary targets could occur on either channel (ear). If automatic processes require little or no capacity, then the processing load induced by monitoring two channels of input versus one channel of input should have little effect on a CM primary task or on a CM secondary task. As stated by Shiffrin and Schneider (for visual tasks), when automaticity develops as a result of CM training, "divided-attention deficits will not be seen because the target stimulus will be detected automatically, in parallel with the other stimuli, and often independently of the other stimuli" (Shiffrin & Schneider, 1977, p. 164).

The predictions for Experiment 2 are as follows. First, if automaticity develops under CM conditions, then performance on the CM primary task should be little affected by a concurrent processing load due to a CM or VM secondary task. Second, if there is no processing cost to the development of automaticity when attention is divided across multiple channels of input, then performance on

the CM primary task and on the CM secondary task should be similar across single- and multiple-channel conditions. However, if CM performance suffers under multiple-channel conditions, this result would suggest that dividing attention across auditory spatial channels may interfere with the development of automaticity.

Method

Subjects. Twelve paid volunteers obtained from the State University of New York at Buffalo student population and the surrounding university community were employed as subjects. All subjects were right-handed, native speakers of English who reported no history of a speech or hearing disorder. Each subject participated in 13 1-h sessions (one session per day). The subjects were paid at the rate of \$4 per hour.

Stimuli. The stimuli consisted of 24 naturally produced stop-consonant-vowel syllables spoken by one male talker and one female talker, for a total of 48 items. The 24 syllables consisted of the same CV combinations used in Experiment 1. Digitization, editing, and all other aspects of creating the stimulus materials were identical to Experiment 1. In a separate pilot experiment, the intelligibility of the male-voice and the female-voice syllables was tested. The results indicated that no significant differences in intelligibility between talkers were present.

Procedure. The experimental factors of secondary task mapping (CM or VM) and channel (single or multiple) were manipulated. Thus, four experimental conditions were created: CM single, CM multiple, VM single, and VM multiple. A constant memory set size of one was used for both primary and secondary target sets. Secondary task was manipulated as a between-subject factor, and channel was manipulated as a within-subject factor. The subjects were equally divided and randomly assigned to two groups of 6 subjects. Each subject participated in 13 total sessions (days), with the first 7 sessions consisting of the one-channel task and the last 6 sessions consisting of the two-channel task. The first session was conducted as practice and was not included in the final data analysis. The number of sessions was determined using a stabilization criterion of non-significant differences in RT over 2 days of training in the CM condition.

The subjects were presented with lists of dichotically presented CV nonsense speech syllables and were required to make appropriate responses whenever a designated target occurred. There were two types of targets: primary and secondary. Each target consisted of a syllable-initial consonant. When performing the monitoring task, the subjects held the primary and secondary targets in memory and made one of two responses each time a target was encountered. Each time a primary target occurred, the subjects were instructed to respond by pushing a button on a response box as quickly as possible. Each time a secondary target occurred, the subjects responded by uttering the word *stop* into a voice-activated microphone as quickly as possible. The use of one manual response and one voice response was designed to eliminate the possibility of response competition arising from making two manual responses (see Nusbaum, 1981). Thus, the subjects monitored the incoming speech syllables and made separate responses to the primary and secondary targets randomly interspersed within each list. The subjects were instructed that their responses to the CM primary targets were of greater importance and priority than their responses to the secondary targets.

The dichotic stimulus pairs were presented with male-voice stimuli in one ear and female-voice stimuli in the other ear. The use of different-voice stimuli in each ear was designed to minimize the possibility of fusions or blends occurring between dichotic stimuli (see Cutting, 1976). Phonetic feature similarity (i.e., voicing, place of articulation) and vowel were varied randomly between the two stimuli comprising each dichotic pair. The target stimuli were ran-

domly interspersed among distractor stimuli within each list. Two different syllable-initial consonants were designated as the primary and secondary targets for each subject, with the constraint that the two target consonants differed on both place and voicing features (e.g., /p/ and /g/).

In the single-channel condition (dichotic presentation, one-channel monitoring), the subjects were presented with primary targets, secondary targets, and distractors in the male-voice ear. Concurrently, distractors were presented in the female-voice ear. The subjects were instructed to monitor the male-voice ear for the target stimuli and to ignore the female-voice ear. In the multiple-channel condition (dichotic presentation, two-channel monitoring), primary targets, secondary targets, and distractors were again presented in the male-voice ear. However, secondary targets and distractors were also presented in the female-voice ear. In this condition, the subjects were instructed that their primary task was to monitor the male-voice ear for the primary targets. However, they were also told to monitor both male- and female-voice ears for the secondary targets. As a result, attention was divided across the two channels (ears) in order to perform the secondary task.

For both VM and CM conditions, *the primary target was always consistently mapped*. That is, the primary target remained exactly the same over all sessions for each subject. Each of the 6 subjects in each mapping condition received a different primary target consonant. In the CM condition, the secondary target was also consistently mapped. In the VM condition, the secondary target was changed from block to block of lists and from session to session. In this condition, the primary target for each subject was paired with one of the five remaining stop consonants to produce the primary-secondary target pairings for each block of lists. For example, one subject had a 't' primary target and a 'g' secondary target for one block of lists, a 't' primary target and a 'k' secondary target in another block of lists, and so on.

The stimuli were presented in lists of 20 dichotic pairs presented every 1,005 msec. Nine lists of pairs occurred in each block, with nine blocks per session. No more than two targets occurred in succession in a list, with no targets occurring in the first or last position in the list. The stimuli in the dichotic pairs were always presented simultaneously, with a target on one ear always paired with a distractor on the other ear. There were 36 primary targets and 24 secondary targets randomly distributed over each block of nine lists. The stimulus presentation order within each list and block and the order of blocks per session were randomized. Ear of presentation of the primary targets was counterbalanced across subjects in both mapping conditions. The primary and secondary targets for each block were continuously displayed on a CRT screen placed directly in front of the subject. The subject initiated each trial by pressing a button. Stimulus presentation and data collection were controlled on-line by a PDP-11/34A computer. RTs were recorded from stimulus onset.

Results and Discussion

The RT data were plotted over sessions for primary and secondary targets. In Figure 2, RT performance over sessions for primary targets is displayed as a function of secondary task mapping and channel condition.

Recall that the first six sessions constituted the single-channel condition and the last six sessions constituted the multiple-channel condition. As shown in Figure 2, RT decreased for the CM primary targets over the first six sessions for both CM and VM secondary target mapping conditions. When the task was changed from single channel to multiple channel on Session 7, a substantial increase in RT was observed for both CM and VM mapping conditions. During the last six sessions, RT to CM primary

targets decreased across sessions when the secondary targets were consistently mapped. However, RT to CM primary targets remained constant and decreased little over training when the secondary targets were mapped in varied fashion. These observations suggest that there is a processing cost in terms of RT to CM primary targets only when a VM secondary task is present *and* multiple channels of input must be monitored.

Figure 3 displays RT performance over sessions for secondary targets as a function of mapping, channel condition, and ear (attended or unattended). In this figure, RT is displayed over 12 sessions for the attended ear and over six sessions for the unattended ear, because secondary targets only occurred in the unattended ear in the multiple-channel condition. For the secondary attended targets, RT decreased over sessions in both CM and VM conditions during the first six sessions. When the task was changed to multiple-channel on Session 7, RT increased for both mapping conditions, although the increase appeared to be greater in the VM condition. During the last six sessions, RT continued to decrease in the CM condition, whereas RT remained fairly constant in the VM condition. It appears that RT performance was detrimentally affected most when the secondary targets were mapped in a varied fashion under multiple-channel conditions. With regard to RT performance over sessions for the un-

attended ear, RT appeared to be slower in the VM condition than in the CM condition and was slower overall relative to the attended secondary targets.

The data were analyzed in terms of overall hits (percent correct), false alarms (percent FAs), and RTs. As in Experiment 1, the data were collapsed over the last two sessions within each experimental condition for further analysis. Two-way ANOVAs were run for the factors of mapping and channel on RT, hits, and false alarms for primary and secondary targets. Performance averaged over the last two sessions (RT, percent correct hits, and percent false alarms) is displayed in Table 2 for primary and secondary targets as a function of mapping and channel conditions.

First, the CM primary target analyses are described. A significant main effect of channel condition was observed for hits [$F(1,10) = 17.3$, $MS_e = 0.0009$, $p < .01$], with fewer hits obtained in the multiple-channel condition. A main effect of channel condition was also found for false alarms [$F(1,10) = 6.6$, $MS_e = 0.00004$, $p < .03$], with more false alarms in the multiple-channel condition. For RT, a nearly significant interaction of mapping with channel condition was obtained [$F(1,10) = 4.6$, $MS_e = 1,246.3$, $p < .06$]. RT (at the end of training) was faster in the multiple-channel condition than in the single-channel condition for the CM group, whereas RT

CM Primary Target RT

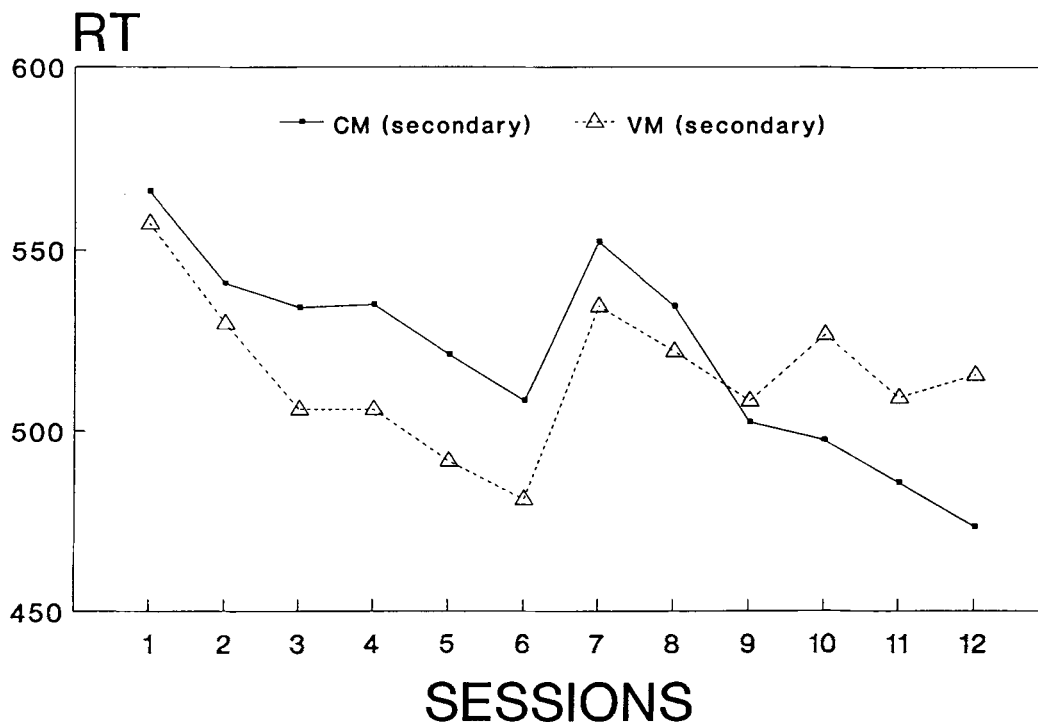


Figure 2. Mean reaction time to primary targets plotted over 12 sessions for the consistent- and varied-mapping groups for Experiment 2. Sessions 1–6 constitute the single-channel condition; Sessions 7–12 constitute the multiple-channel condition.

Secondary Target RT

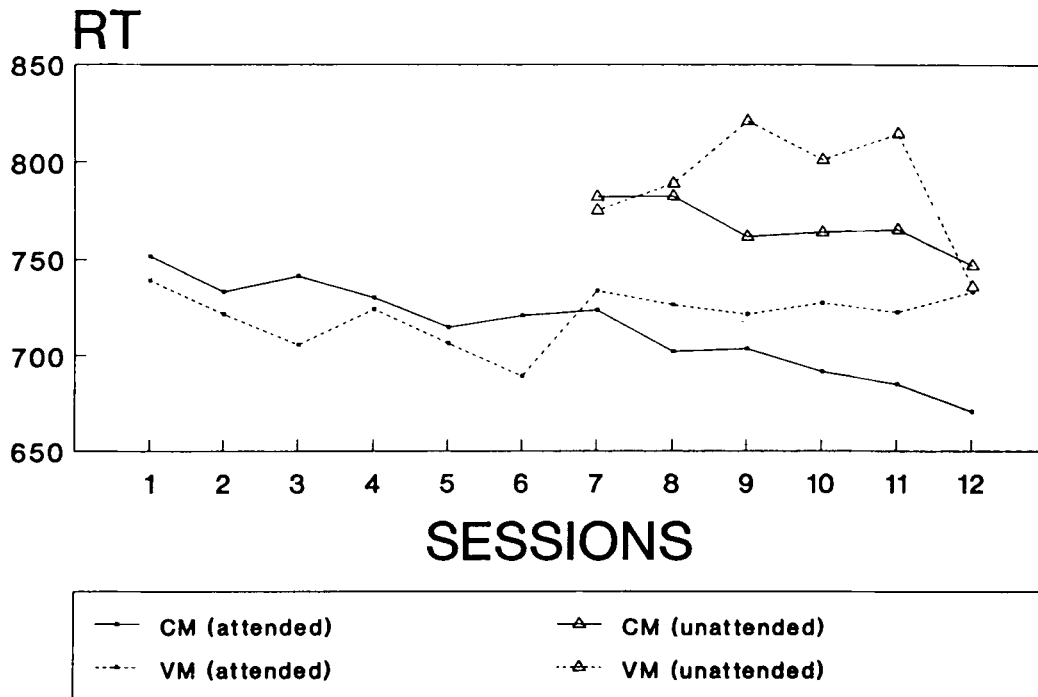


Figure 3. Mean reaction time (RT) to secondary targets plotted over 12 sessions for the consistent- and varied-mapping groups for Experiment 2. Sessions 1–6 constitute the single-channel condition; Sessions 7–12 constitute the multiple-channel condition. RTs are also plotted for the unattended ear for consistent and varied mapping for the multiple-channel condition.

was slower in the multiple-channel condition than in the single-channel condition for the VM group. To summarize, less hits and more false alarms were made to CM primary targets in the multiple-channel condition, and RT to CM primary targets was slower under VM multiple-channel conditions.

Next, the data for secondary targets occurring on the attended ear (i.e., the ear that the primary targets occurred on) are described. A significant main effect of mapping was observed for hits [$F(1,10) = 13.7$, $MS_e = 0.006$, $p < .05$], with fewer hits in the VM condition. A main effect of channel condition was also obtained for hits [$F(1,10) = 13.3$, $MS_e = 0.0024$, $p < .01$], with fewer hits in the multiple-channel condition. A significant main effect of channel condition was also found for false alarms [$F(1,10) = 11.8$, $MS_e = 0.0003$, $p < .01$], with more false alarms in the multiple-channel condition. A significant interaction of mapping with channel condition was observed for hits [$F(1,10) = 5.56$, $MS_e = 0.0024$, $p < .04$]. Post hoc tests of the interaction revealed that there were significantly fewer hits in the VM multiple condition than in the other three conditions. A significant interaction of mapping with channel condition was also obtained for RT [$F(1,10) = 5.57$, $MS_e = 1,312.3$, $p < .04$]. Post hoc tests revealed significant differences in RT

between VM multiple and CM multiple conditions only. To summarize, there were fewer correct responses and more false alarms in the multiple-channel condition, and fewer correct responses in the VM condition. The interaction of mapping \times channel condition for hits indicated that performance for VM targets in the multiple-channel condition was worst. The interaction of mapping \times channel condition for RT was similar to the nearly significant interaction observed for the CM primary target data.

Finally, analyses comparing secondary target performance on the attended ear (the ear in which the primary targets occurred) with secondary target performance on the unattended ear (the ear in which primary targets did not occur) are described. Separate two-way ANOVAs for RT and hits were performed with the factors of mapping and ear (attended or unattended). Analyses of the false-alarm data were not conducted.² A significant main effect of ear was observed for RT [$F(1,10) = 59.1$, $MS_e = 400.5$, $p < .01$], with RT slower for the unattended ear. A significant main effect of ear was also obtained for hits [$F(1,10) = 93.2$, $MS_e = 0.0144$, $p < .01$], with fewer hits for the unattended ear. Post hoc tests indicated that these effects were significant for both mapping conditions. A significant main effect of mapping was also observed for hits [$F(1,10) = 33.6$, $MS_e = 0.011$, $p < .01$], with

Table 2
Mean Reaction Times (in Milliseconds), Hits (% Correct), and False Alarms (% FAs) for Responses to Primary Targets, Secondary Targets Occurring in the Attended Ear, and Secondary Targets Occurring in the Unattended Ear as a Function of Mapping and Channel Conditions for Experiment 2

Condition	CM Group			VM Group		
	RT	Hits	FAs	RT	Hits	FAs
Primary Targets						
Single-channel	514.8	96.8	0.6	486.1	97.8	0.7
Multiple-channel	479.4	93.7	1.7	512.5	90.0	1.1
Secondary Targets						
Single-channel (Attended Ear)	717.6	94.5	0.4	697.7	87.8	0.6
Multiple-channel (Attended Ear)	677.9	91.9	3.8	727.7	75.8	1.7
Multiple-channel (Unattended Ear)	756.1	53.4		775.1	19.7	

Note—CM = consistent mapping; VM = varied mapping.

fewer hits for the VM condition. To summarize the results of comparing secondary target performance on the attended ear to the unattended ear, slower and less accurate responses to secondary targets were observed for targets occurring on the unattended ear for both mapping conditions. In addition, fewer correct responses to secondary targets were obtained when the stimuli were mapped in a varied fashion.

Although the results of Experiment 2 are complex, three important results stand out. First, recall that the primary targets were always consistently mapped. Yet, when attention was divided across multiple input channels, responses to the CM primary targets became significantly less accurate. Second, when the secondary targets were variably mapped, latencies to CM primary targets were slower in the multiple-channel condition than in the single-channel condition. Third, when attention was divided across channels for CM secondary targets, responses to secondary targets occurring in the unattended channel were significantly slower and less accurate than responses to secondary targets in the attended channel. These three results (that accuracy to CM primary targets was affected by dividing attention over spatial channels, that latencies to CM primary targets were slower when dividing attention over channels while performing a concurrent VM task, and that responses to CM secondary targets were worse on the unattended channel) indicate that the development of automaticity to speech syllables under CM conditions was significantly affected by dividing attention across multiple channels of speech input. Although processing was faster and more accurate for CM training when only a single channel is considered, we have demonstrated that there is a processing cost to the development of automatic responses to speech stimuli when multiple channels must be monitored, especially under concurrent VM task load conditions.

GENERAL DISCUSSION

The results of the present study provide information pertinent to the development of automaticity in auditory/speech processing. First of all, the results of Experiment 1

indicate that after extended CM training, perception is relatively fast and relatively unaffected by increases in short-term memory load. In contrast, perception under VM conditions is slower and is substantially interfered with by memory load. This finding indicates that, under single-channel conditions, automatic attentional responses to speech can be learned through CM training and can ultimately facilitate perceptual processing. This result is similar to findings using visually based paradigms, and it extends the findings of Poltrock et al. (1982) to single-channel monitoring conditions.

However, when performance is examined under multiple-channel conditions with concurrent dual-task processing loads, we find that the development of automaticity is interfered with. In Experiment 2, responses to CM primary targets were less accurate when attention was divided over two spatial input channels, and latencies to CM primary targets were slower when a concurrent VM task was performed under multiple-channel conditions. Also, responses to CM secondary targets differed across attended and unattended channels in the multiple-channel condition, a result that should not occur if the processing of secondary targets across channels is accomplished in parallel without cost. So, when multiple channels of input are monitored, performance under CM training conditions is disrupted and results in a processing cost, relative to performance under single-channel conditions.

The results of Experiment 2 contrast with the results of visually based tasks, indicating that concurrent processing loads do not affect consistently mapped tasks (Fisk & Schneider, 1983; Schneider & Fisk, 1982a; Strayer & Kramer, 1990). These results also conflict with data for auditory stimuli cited by Moray (1975), in which performance under extended practice conditions is similar across single- and multiple-channel environments. Why do our results differ from this previous work? With respect to the visual studies, the obvious difference is that we are examining processing over two auditory spatial channels of input. This situation may induce an attentional load that has a qualitatively different effect on the development of automaticity than does memory load or frame size in visually based tasks. With respect to the auditory studies,

it should be noted that Moray (1975) did not explicitly manipulate mapping in terms of CM and VM tasks and that comparisons of binaural (single-channel) and dichotic presentation conditions were conducted without manipulations of concurrent processing load. In addition, Poltrock et al. (1982) did not compare single- and multiple-channel performance. It is possible that the tasks used in both the visual and the auditory studies did not "stress" subjects' processing capacity to the extent that one could exhibit perceptual interference, such as we obtained.

Given the results we obtained in the present study, what conclusions can be drawn about automaticity in the auditory domain? It can be concluded that perceptual responses to speech on a single channel are generally facilitated after training under CM conditions, relative to VM conditions. However, this result is tempered by the finding that, even after substantial training under CM training conditions, processing is interfered with when attention is divided over multiple channels of input. Does this mean that automaticity cannot develop properly under multiple-channel conditions? Not necessarily. It is possible that with extended practice for a much greater number of sessions than we used, performance under multiple-channel conditions would eventually reach performance observed under single-channel conditions. But, the point is that, at least initially, the processing load induced by dividing attention over multiple input channels does interfere with the development of automaticity. This finding suggests that the processes underlying the development of automaticity across visual and auditory modalities may be somewhat different.

Another issue raised by the present results is whether the detrimental effects of monitoring multiple input channels were due to attention "switching" back and forth between spatial input channels, or whether a generic demand for time-shared processing capacity or resources was responsible. With the present results, we are unable to distinguish between these two alternatives. If, indeed, it turns out that attention must be switched back and forth between input channels, with perceptual analysis occurring one channel at a time, this switching may be a "controlled" process that in and of itself requires processing capacity and/or resources that impact the development of automaticity.

In conclusion, the general implication of our results is that we should investigate more carefully the capacity requirements of visual and auditory processing under certain training conditions. It is possible that our results reflect a basic difference in processing for auditory-based versus visually-based stimuli. But, perhaps our results differ from previous studies because the experimental paradigms were not quite sensitive enough to pick up the capacity requirements of processing under CM training conditions (Fisk & Schneider, 1983; Schneider & Fisk, 1982a; Strayer & Kramer, 1990). Future research needs to examine the capacity requirements of automatic processing across perceptual modalities in further detail and under a variety of experimental conditions.

REFERENCES

- BOOKBINDER, J., & OSMAN, E. (1979). Attentional strategies in dichotic listening. *Memory & Cognition*, 7, 511-520.
- CUTTING, J. E. (1976). Auditory and linguistic processes in speech perception: Inferences from six fusions in dichotic listening. *Psychological Review*, 83, 114-140.
- FISK, A. D., & SCHNEIDER, W. (1983). Category and word search: Generalizing search principles to complex processing. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 9, 177-197.
- LABERGE, D. (1975). Acquisition of automatic processing of perceptual and associative learning. In P. M. A. Rabbitt & S. Dornic (Eds.), *Attention and performance V* (pp. 50-64). New York: Academic Press.
- LABERGE, D. (1981). Automatic information processing: A review. In J. B. Long & A. D. Baddeley (Eds.), *Attention and performance IX* (pp. 173-186). Hillsdale, NJ: Erlbaum.
- LOGAN, G. D. (1978). Attention in character-classification tasks: Evidence for the automaticity of component stages. *Journal of Experimental Psychology: General*, 107, 32-63.
- LOGAN, G. D. (1979). On the use of a concurrent memory load to measure attention and automaticity. *Journal of Experimental Psychology: Human Perception & Performance*, 5, 189-207.
- MORAY, N. (1975). A data base for theories of selective listening. In P. M. A. Rabbitt & S. Dornic (Eds.), *Attention and performance V* (pp. 119-136). New York: Academic Press.
- NEUMANN, O. (1984). Automatic processing: A review of recent findings and a plea for an old theory. In W. Prinz & A. F. Sanders (Eds.), *Cognition and motor processes* (pp. 255-293). Berlin: Springer-Verlag.
- NUSBAUM, H. C. (1981). *Capacity limitations in phoneme perception*. Unpublished doctoral dissertation, State University of New York at Buffalo.
- NUSBAUM, H. C., & SCHWAB, E. C. (1986). The role of attention and active processing in speech perception. In E. C. Schwab & H. C. Nusbaum (Eds.), *Perception of speech and visual form: Theoretical issues, models, and research* (pp. 113-157). New York: Academic Press.
- OSTRY, D., MORAY, N., & MARKS, G. (1976). Attention, practice, and semantic targets. *Journal of Experimental Psychology: Human Perception & Performance*, 2, 326-336.
- POLTROCK, S. E., LANSMAN, M. L., & HUNT, E. (1982). Automatic and controlled processes in auditory target detection. *Journal of Experimental Psychology: Human Perception & Performance*, 8, 37-45.
- POSNER, M. I., & SNYDER, C. R. R. (1975). Facilitation and inhibition in the processing of signals. In P. M. A. Rabbitt & S. Dornic (Eds.), *Attention and performance V* (pp. 669-682). New York: Academic Press.
- SCHNEIDER, W. (1985). Toward a model of attention and the development of automatic processing. In M. I. Posner & O. S. M. Marin (Eds.), *Attention and performance XI* (pp. 475-492). Hillsdale, NJ: Erlbaum.
- SCHNEIDER, W., & DETWEILER, M. (1988). The role of practice in dual-task performance: Toward workload modelling in a connectionist/control architecture. *Human Factors*, 30, 539-566.
- SCHNEIDER, W., & FISK, A. D. (1982a). Concurrent automatic and controlled visual search: Can processing occur without resource cost? *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 4, 261-278.
- SCHNEIDER, W., & FISK, A. D. (1982b). Degree of consistent training: Improvements in search performance and automatic process development. *Perception & Psychophysics*, 31, 160-168.
- SCHNEIDER, W., & SHIFFRIN, R. M. (1977). Controlled and automatic human information processing: I. Detection, search, and attention. *Psychological Review*, 84, 1-66.
- SHIFFRIN, R. M., & SCHNEIDER, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending, and a general theory. *Psychological Review*, 84, 127-190.
- STERNBERG, S. (1966). High-speed scanning in human memory. *Science*, 153, 652-654.
- STRAYER, D. L., & KRAMER, A. F. (1990). Attentional requirements of automatic and controlled processing. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 16, 67-82.

NOTES

1. Several aspects of our practice schedule differed from typical research using this paradigm (Schneider & Shiffrin, 1977; Poltrock et al., 1982). First, target set size was a between-subject variable. We manipulated set size between subjects and the mapping variable within subjects because we felt that if between-group variability had any obscuring effect on the results, set size would be affected less. Second, the number of sessions for one-target training versus three-target training differed. This was due to the fact that we adopted a "stabilization criterion" based on nonsignificant differences in RT for 2 consecutive days of CM training. We felt that using an arbitrary and equal number of training sessions for the different target set size conditions would be problematic. For large set size conditions, it is likely that more practice is needed initially for subjects to learn how to automate processing than under low memory-load conditions. Comparisons across an arbitrary and equal number of training sessions may be misleading, because differences across set sizes at the end of training may simply reflect this initial disparity. By using an operationally defined criterion, as we did, to compare performance across set sizes when additional practice has little effect, the residual effects of memory load across mapping condi-

tions can be properly assessed. Third, mapping was manipulated as a within-subject variable. This means that subjects receiving CM training before VM training may encounter a problem of "negative transfer." That is, CM targets that become distractors in VM training may automatically attract attention to them, resulting in worse performance and inflated RT values in the VM conditions. However, we used a reasonably large number of training sessions across set size conditions (10 and 15), so we expected that any interfering responses of this sort would quickly extinguish. In addition, if negative transfer existed, one would expect the false-alarm rate to increase substantially. Thus, inspection of the false-alarm data provides some indication of whether negative transfer may have occurred. And fourth, our target/distractor set was relatively small (six phonemes) compared with that in other studies. Fewer distractors may make the overall task easier for both CM and VM conditions.

2. Since false alarms to both attended-ear and unattended-ear secondary targets were recorded by virtue of the same voice response, partitioning of the secondary target false-alarm data over ears was not possible.

(Manuscript received September 12, 1990;
revision accepted for publication July 9, 1991.)

Notices and Announcements



SIXTH ANNUAL EDUCOM HIGHER EDUCATION SOFTWARE AWARDS COMPETITION

DISCIPLINES FOR THE 1992 COMPETITION

- NATURAL SCIENCES
- SOCIAL SCIENCES
- ACCOUNTING

Deadline for the
1992 Competition is
February 24, 1992

The EDUCOM Higher Education Software Awards Program was established in 1987 to recognize and promote quality educational software and computer-based teaching innovations in higher education. The program has two divisions:

- Product Division, for original software.
- Curriculum Innovation Division, for the innovative uses of new or existing software to solve important instructional problems.

Entry Forms: For entry forms and further information please contact:

Higher Education Software Awards Program, attn: Gail Miller
Computer Science Center, Building 224
University of Maryland
College Park, Maryland 20742-2411
(301) 405-7534 E-mail: awards@crystal.umd.edu

Charter Sponsors:

Apple Computer, Inc. • IBM Corp. • NeXT Computer

Contributing Sponsors:

Bell Atlantic Corp., Claris Corp., Microsoft Corp., National Center for Automated Information Retrieval (NCAIR), Sun Microsystems.

Other Sponsors: EDUCOM, The University of Maryland