Memory & Cognition
1993, 21 (3), 318-328

Disruption of visual short-term memory by changing-state auditory stimuli: The role of segmentation

DYLAN M. JONES, WILLIAM J. MACKEN, and ALISON C. MURRAY
University of Wales College of Cardiff, Cardiff, Wales

Typically, serial recall performance can be disrupted by the presence of an irrelevant stream of background auditory stimulation, but only if the background stream changes over time (the auditory changing-state effect). It was hypothesized that segmentation of the auditory stream is necessary for changing state to be signified. In Experiment 1, continuous random pitch glides failed to disrupt serial recall, but glides interrupted regularly by silence brought about the usual auditory changing-state effect. In Experiment 2, a physically continuous stream of synthesized vowel sounds was found to have disruptive effects. In Experiment 3, the technique of auditory induction showed that preattentive organization rather than critical features of the sound could account for the disruption by glides. With pitch glides, silence plays a preeminent role in the temporal segmentation of the sound stream, but speech contains correlated time-varying changes in frequency and amplitude that make silent intervals superfluous.

Irrelevant background speech disrupts serial short-term memory for material presented visually (see D. M. Jones & Morris, 1992, in press, for reviews). A reasonably clear picture of the features of the task and of the speech that elicits this effect is now emerging, as are the general implications for our understanding of attentional factors in memory (see D. M. Jones & Broadbent, 1991). Two views of this disruption are contrasted here. One is that the extent of disruption depends on the degree to which the sound resembles speech. Salamé and Baddeley (1989) have proposed that the effects of speech and nonspeech sounds on serial recall could be understood in terms of a filter or detector system that selectively passes into memory sounds that resemble speech. We wish to put forward an alternative model, the *changing-state hypothesis*, whose pivotal feature is that the sound has to show particular variation over its time course (see D. M. Jones, in press). Our primary concern in the present paper is to refine the notion of change within the auditory changing-state concept; our secondary purpose is to assess the view that sound has to be "speech-like" before it disrupts serial recall. The purpose of the experimental work reported here is to identify cues common to speech and nonspeech stimuli that nevertheless may serve as the basis for disruption. Specifically, interest is attached to the role of segmentation in signifying changing state, and to the ways this may lead to disruption of recall by both speech and nonspeech stimuli. In the present context, we use the term *segmentation* to describe the detection of boundaries based on analysis at a physical level, and not, as is commonly the case, to denote boundaries that have some linguistic connotation.

Evidence so far collected suggests that when irrelevant streams consist of strings of discrete elements (utterances or tones), the changes between the consecutive elements that comprise a stream seem to give it its disruptive qualities (D. M. Jones, in press; D. M. Jones, Madden, & Miles, 1992; Morris & D. M. Jones, 1990a, 1990b). But it is not yet clear whether changing state may be manifested only with discrete stimuli (that is, stimuli separated by silence) or whether it may arise from any stimulus that has a discriminable change in state, including continuous stimuli. That a hummed tune fails to disrupt serial recall but sung speech produces the usual degree of disruption (Morris, Quayle, and D. M. Jones, 1987) suggests that not all sounds produced by the human voice are equally potent disruptors and that the ease with which sound may be segmented into its component parts accounts for the degree of disruption. This indirect evidence is complemented by other instances in which continuous but varying stimuli have been found to produce little disruption of serial recall; these include amplitude-modulated noise (D. M. Jones et al., 1992; Salamé and Baddeley, 1989) and an amplitude-modulated continuous tone (D. M. Jones et al., 1992). In the experiments reported here, we explored whether continuous but varying sounds bring about the changing-state effect, as well as the degree to which the segmentation of such sounds is a necessary precondition for disruption by changing state.

The changing state hypothesis underpins a more general model of working memory that is based on a black-

Thanks are due Richard Blight for his general help in setting up the equipment. Helen Bowman and Craig Steel ran some of the subjects. The work was supported by a research grant from the U.K.'s Economic and Social Research Council to the first author. Correspondence should be addressed to D. Jones, School of Psychology, University of Wales College of Cardiff, P.O. Box 901, Cardiff CF1 3YG, Wales, United Kingdom.

board analogy called the object-oriented episodic record (O-OER) model (D. M. Jones, in press). A central feature of this model is the formulation of "objects" on an episodic surface. We regard the process of segmentation of the auditory stream as central to the formation of objects: these objects are first discerned by the detection of acoustic cues that indicate boundaries to events; some judgment is then made about whether successive objects are different, in which case they occupy unique locations on the surface. Objects on the episodic surface are joined by pointers that reflect the organization of streams of information from auditory and visual sources. The process of interference with serial recall is explained by the coexistence in short-term memory of objects from deliberate subvocal rehearsal of the visual material and objects of auditory origin derived from the irrelevant sound stream. Disruption occurs as a result of a conflict in organization of two sets of pointers, one from material that is deliberately rehearsed and the other from the streams set down preattentively by auditory perceptual processes. If an item is repeated in the auditory channel, the pointers are said to be "self-referencing" and point only to themselves. In this case, the competition between pointers is minimized, which in turn accounts for the minimal disruptive effect on serial recall caused by irrelevant streams consisting of one repeated item. We therefore expect the disruption by irrelevant speech to be confined to serial recall and absent from tasks that demand free recall (D. M. Jones & Macken, 1993; Salamé & Baddeley, 1990).

A major competing interpretation of the irrelevant speech effect is that some filter or detector that passes speech but not nonspeech signals into short-term memory is responsible for the irrelevant speech effect (Salamé & Baddeley, 1982, 1989). On the basis of this type of model, one might expect the degree of disruption to be in proportion to the similarity of the irrelevant sound to speech, but the evidence is equivocal. For example, the theory is weakened by several lines of evidence that suggest that speech is neither a sufficient condition (D. M. Jones et al., 1992) nor a necessary condition (D. M. Jones & Macken, 1993) for disruption of serial recall. Salamé and Baddeley (1989) have proposed that the capacity for *music* to disrupt recall is related to the degree to which the sound resembles speech or contains speech. Although they showed that the effect of narrative speech on serial recall was more pronounced than that of either instrumental or vocal music, a result that concords with this view, they also found that vocal music was no more disruptive than instrumental music, which is not expected on this view. They concluded that although the trend was in the correct direction, their experiments showed "no solid evidence for the differences between the various vocal and instrumental excerpts" (Salamé & Baddeley, 1989, p. 119). It is not easy to refute the theory of Salamé and Baddeley, since they do not specify what parameters of a sound distinguish speech from nonspeech. Indeed this distinction is the subject of long-standing controversy (a substantial literature exists on this topic; for a convenient summary see Samuel & Tartter, 1986). A more useful analytic method is to specify the acoustic characteristics that determine disruption by irrelevant auditory stimuli. Such an approach would be particularly powerful if the characteristics were suggested a priori by a theory. By specifying that object formation is a necessary condition for disruption of serial recall, the O-OER model points to the importance of acoustic factors that segment the acoustic stream.

In the three experiments reported here, we dealt with two propositions derived from the changing-state hypothesis that suggest the means by which segmentation will take place. In the first two experiments, we were concerned with the proposition that perceptual discontinuity is a necessary precondition for bringing about disruption of serial recall and that the use of silence during the stream is one of the ways in which this may be achieved. In the third experiment, we were concerned with establishing the extent to which organizational factors determine the degree of disruption; we also intended to test the model of Salamé and Baddeley (1989).

EXPERIMENT 1

In Experiment 1, the role of segmentation in nonspeech stimuli was explored by contrasting the action of two types of auditory stream that are made up of random frequency glides (glissandi). The effect of continuous glides was contrasted with glides regularly interrupted by silence. The basic material for the experiment was created by generating a continuous sound in which the only source of variation is frequency. By this means, we attempted to construct a more carefully controlled analogue of the humming material that had previously been shown to be ineffectual in disrupting serial recall (see Morris et al., 1987). The same glides, this time regularly interrupted by short periods of silence, were used as the basis for the second type of auditory material. We made the strong prediction that the continuous glissandi would have little or no effect on the accuracy of serial recall but that the interrupted glissandi would have a marked disruptive effect on serial recall, just as we had found for steady-state tones separated by short periods of silence and for speech (D. M. Jones & Macken, 1993).

The short-term memory task employed here is similar to that employed by Jones et al. (1992). In this task, seven consonants are presented visually for serial recall, followed by a 10-sec interval in which the subject is expected to rehearse covertly before being cued for a written response. When dictated by the experimental design, auditory material is presented during presentation and rehearsal of the list. (Miles, D. M. Jones, & Madden, 1991, have shown that irrelevant speech has its effect at both presentation and rehearsal stages of the task.)

Method

Subjects. Twenty volunteers, drawn from undergraduates and postgraduates at the University of Wales College of Cardiff were

paid an honorarium for taking part in the study. All reported normal hearing.

Apparatus and Materials. Items to be recalled were presented serially on the screen of a Macintosh LC microcomputer via HyperCard software. Seven uppercase consonants, F, K, L, M, Q, R, and Y, were presented in a different randomized order on each trial.

Sound was delivered via Beyer electrostatic headphones (Type ET 1000) from recording of the sounds held in digital format as SND resources within the Macintosh. Two types of auditory material were produced: continuous glissandi and interrupted glissandi. These are illustrated schematically in Figure 1. The top panel shows how the frequency of the sound varies over time in the continuous glissandi. This variation is random within a restricted range (less than 0.7 Hz). The lower panel illustrates the variation in frequency of interrupted glissandi, using the same pattern of pitch change but this time with regular interruption by silence (signified by the solid vertical bars).

Continuous glissandi were generated by using low-pass filtered noise to drive a voltage-controlled oscillator. Pink noise from a Consilium Industri noise generator (Model PNG 11) was low-pass filtered by a Barr and Stroud (Type EF3) filter set to pass frequencies below 0.7 Hz with a roll off in attenuation of 24 dB/octave above that point. The resulting signal was amplified by 20 dB with an amplifier of our own construction (giving linear response to signals from dc to ultrasound). This amplified, randomly varying signal served as a control voltage for a Farnell (Model DSG1) signal generator acting as a voltage-controlled oscillator. As the voltage varied in a random fashion, so did the pitch of the sound produced by the oscillator. The *range* over which the pitch varied (the depth of modulation) was adjusted to give a discernible degree of variability, while at the same time ensuring that the excursions of pitch at the lower frequency bound did not lead to the perception of silence that is due to the insensitivity of the ear to low frequencies. The low-pass frequency and the range of the glide were set in pilot trials. Two considerations guided their selection: that the overall temporal variability should be judged similar to that of speech and that the glide did not contain periods of silence. As a result of the pilot work, the glide was set to cover the range of 350–950 Hz.

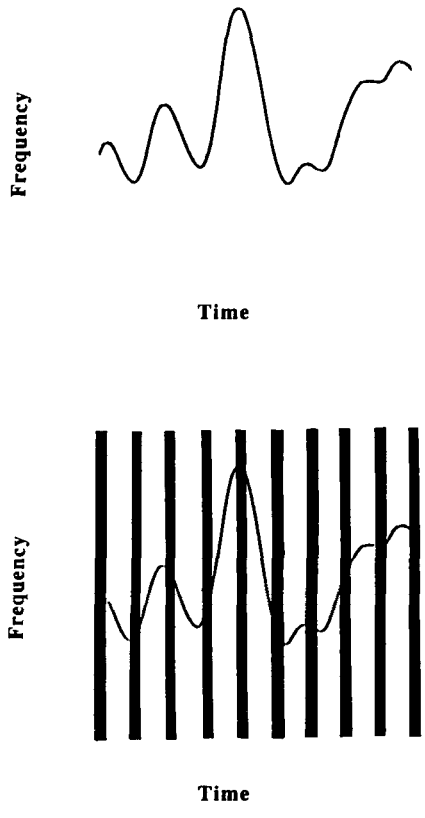

Figure 1. Schematic diagram of the material used in Experiment 1. The upper panel shows the pattern of variation of frequency over time produced by low-pass filtered noise that was used to drive a voltage-controlled oscillator. The lower panel shows the same pattern of frequency change, but this time the glides are periodically interrupted by silence.

Interrupted glissandi were produced by the same equipment, but with the addition of an electronic audio switch of our own construction that served to produce regular periods of silence. The rise and fall time of the switch was designed to minimize the possibility that audible clicks would be generated as a side effect of its activity. The action of the switch was governed by pulses from a Global Specialities Corporation pulse generator (Model 4001). In this way, the glide was interrupted every 300 msec by 200 msec of silence. The selection of these values (in conjunction with those of the frequency of oscillation governed by the setting of the low-pass filter) gave rise to stimuli that contained discernible and continuous pitch variation.

The resulting stimuli were digitized to 8-bit resolution at a sampling rate of 22 kHz (using MacRecorder software by Farallon) and edited to provide excerpts that were stored as SND resources for use within a Macintosh HyperCard environment.

Design. Trials representing each of the conditions were presented quasirandomly, with the constraint that each condition would be presented before any condition was repeated—quiet, continuous glissandi and interrupted glissandi. There were 60 trials in all, 20 for each of the experimental treatments.

Procedure. Subjects were tested individually. Each was seated in a soundproofed and air-conditioned room at a distance of about 0.5 m from the screen. The subject initiated the presentation of each list by selecting a HyperCard button with a mouse-driven pointer, and the consonants were presented individually at a rate of one per second (on for 800 msec, off for 200 msec). When seven consonants had been displayed, the word "wait" was flashed on and off for 10 sec. During the 10-sec delay, the subject was expected to rehearse covertly. The word "recall" was then displayed, to prompt the subject to make the response. When the written response was complete, the subject used the mouse to initiate another trial. The irrelevant sound was played over the headphones during presentation and rehearsal, and it was switched off automatically during recall. The experiment was preceded by a short practice session in control conditions. The subject was given standard written instructions on the computer screen. The instructions emphasized that one should ignore the sound that one heard, that it would not contain any messages, and that the subject would not be tested on its contents. The experiment took some 40 min in all.

Results and Discussion

Performance was scored with respect to serial position. Error data were subjected to a 3 (auditory condition) × 7 (serial position) analysis of variance (ANOVA) that showed main effects of condition [$F(2,38) = 3.49$, $MS_e = 11.52$, $p < .04$] and serial position [$F(6,114) = 6.79$, $MS_e = 12.67$, $p < .0001$]. There was also a significant interaction between the two factors [$F(12,288) = 2.46$, $MS_e = 1.40$, $p < .005$]. As can be seen in Figure 2, in which the number of correct responses is plotted, this interaction is due to the absence of an effect of auditory condition in the primacy portion of the serial position curve. We do not attach any theoretical importance to the locus of the serial position effect.

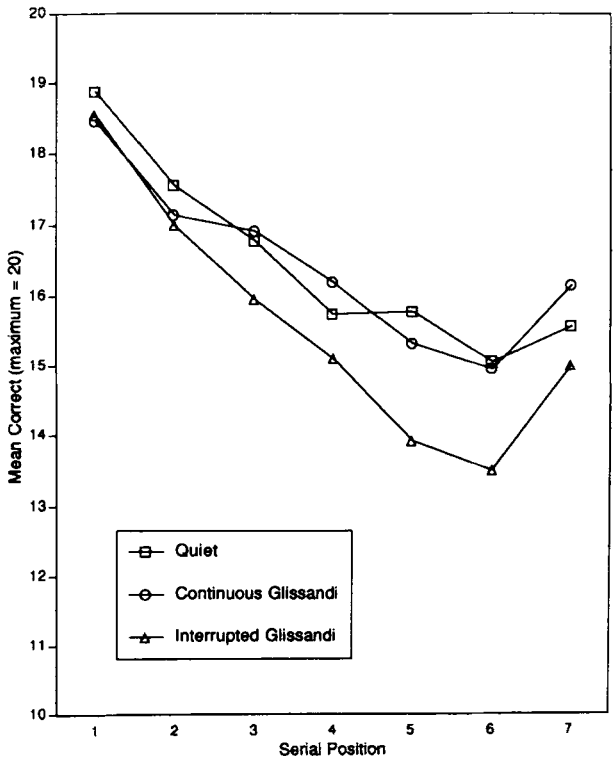

Figure 2. Mean correct responses for each serial position in Experiment 1, contrasting quiet, interrupted glissandi, and continuous glissandi. Scores are out of a maximum of 20.

Multiple comparisons revealed greater disruption for interrupted glissandi than for quiet ($p < .05$) and continuous glissandi ($p < .02$). The difference between the continuous and quiet glissandi conditions did not reach significance ($p > .05$).

The outcome of the experiment is unequivocal: only interrupted glissandi produce appreciable disruption. This points to the fact that within the range of variability studied here, changing state is signified by discontinuous (varying) stimuli rather than by continuous varying stimuli. The important contrast between the results reported here and those previously obtained with nonspeech sounds is that the current result directly addresses the issue of why only certain nonspeech sounds have been shown to produce disruption of serial recall. For example, it may explain why certain kinds of music would be more disruptive than others: only streams that have cues to segmentation (that may be present in speech or nonspeech streams) will have the power to disrupt. In addition, the results echo those found in the contrast between humming and singing shown by Morris et al. (1987), who argue that humming and singing differ in the degree to which cues to segmentation are present in the signal; but out of necessity, the degree of segmentation was not closely controlled and their result could have arisen from one or more of a range of confounding factors. The results of the present Experiment 1 counter any doubts about that outcome. They also imply that the disruption produced by music will not only be brought about by sung lyrics but may equally be produced by particular types of instrumental passages. In turn, this means that unless care is taken to match music excerpts on some criterion such as the frequency with which the stream can be segmented, the effect of music on performance will be rather inconsistent. It was not our intention to explore this point further in the present experiments. In our view, it was more appropriate to focus on acoustic factors that are responsible for the disruption by examining very carefully controlled artificial stimuli rather than face the difficulty of using naturally occurring musical stimuli.

Experiment 2 was designed to test whether segmentation by silence is an important factor when speech is the irrelevant auditory material.

EXPERIMENT 2

Our central hypothesis was that to meet the conditions for changing state, a stream must contain cues that cause it to be segmented by the auditory system. We proposed that intrinsic cues to segmentation are detected automatically for all signals—speech and nonspeech—at a preattentive level (see below for a further discussion, and see Experiment 3 in particular). Nonspeech streams are also subject to such perceptual processes. Experiment 1 served as an illustration that one such cue is silence, but this does not exclude the possibility that a continuous pitch glide could nevertheless be segmented perceptually in terms of changes in energy along its course, as long as these changes in energy are sufficiently sharp. Given this interpretation, the introduction of *silence* into a stream has no general significance, for it is only one of several possible ways of providing a sufficiently sharp transition for marking boundaries between events. A stream of *narrative speech* possesses such attributes as part of its naturally occurring acoustic qualities, and hence it becomes segmented by the perceptual system whether or not it contains periods of silence. Rather, the signification of changing state (by, e.g., changes in energy at syllable boundaries) should be processed preattentively to cue separate events and thereby lead to segmentation.

The rationale for this experiment is an extension of that used in the first of the series: a continuous auditory signal contrasted with one that is regularly interrupted by silence, and with a quiet condition. Because of the strong supposition that perceptual factors in the preattentive segmentation of sound would have a powerful influence, we made the prediction that, unlike the glissandi (which varied only in frequency), the interrupted and uninterrupted speech material would produce equivalent degrees of disruption.

Method

Subjects. Twenty subjects drawn from the same population as those in the previous experiment were paid for their participation.

Apparatus and Materials. Two sequences to be used as irrelevant background streams were created on a parallel formant syn-

thesizer using synthesis by rule (Holmes, 1985). A Loughborough Sound Images Research Synthesizer was driven by parameters created on a BBC Microcomputer running Syncon software (Holmes, 1986). The synthesized material consisted of two types of sequences of vowel sounds. The first type of sequence was continuous: the duration of each vowel was 500 msec, and the formant transitions between vowels were calculated by the software (Holmes, 1986; Holmes, Mattingly, & Shearme, 1964). The second type of sequence had the same vowel sounds, but with the duration reduced to 400 msec and with 100 msec of silence inserted between vowels. The vowels in this second stream were steady state throughout, the silences obviating any need for transitions. The formant frequencies and other specifications necessary for calculation of the parameter values to drive the synthesizer were derived from tables provided as part of the Syncon software that were based on measures of Received Pronunciation English /i/, /I/, /e/, /ɜ/, /a/, /ɔ/, /ʊ/, /u/. The fundamental frequency was kept constant at 104 Hz. Spectrograms for each type of material (produced by Signalyze software; see Keller, 1990) are shown in Figure 3. Approximately 20 sec of each of these sequences was recorded at 48 kHz onto digital audio tape, and then transferred to a Macintosh IIfx microcomputer at a sampling rate of 22 kHz to be stored as SND resources for use within a HyperCard environment.

Procedure. The procedure used here was the same as that used in the previous experiment.

Results and Discussion

A 3 (auditory condition) × 7 (serial position) ANOVA carried out on the error scores yielded significant main effects of both auditory condition [$F(2,38) = 7.35$, $MS_e = 12.33$, $p < .002$] and serial position [$F(6,114) = 7.32$, $MS_e = 16.29$, $p < .0001$], as well as a significant interaction between the two factors [$F(12,228) = 1.89$, $MS_e = 1.97$, $p < .05$]. As can be seen from Figure 4, in which correct responses are plotted against serial position, this interaction is due to the absence of an effect of auditory condition at the first few serial positions. The interaction may be due to a variety of factors, and it is not our intention to explain it here. Planned comparisons revealed that performance in both the continuous and the interrupted voice conditions was significantly worse than in quiet ($p < .0006$ and $p < .007$, respectively) but they did not differ significantly from each other ($p > .05$).

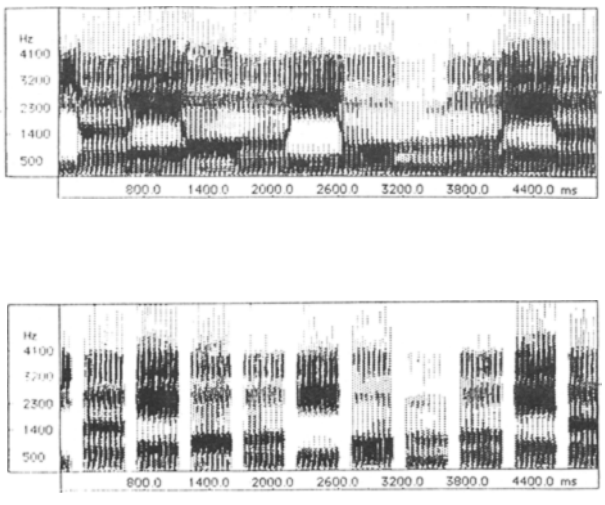

Figure 3. Spectrographs of the stimuli used in Experiment 2. Time is on the abscissa, and frequency is on the ordinate; intensity is coded by shading (with dark being more intense). The upper panel shows a continuously voiced stimulus (fundamental, 104 Hz); the lower panel shows the speech interrupted by silence.

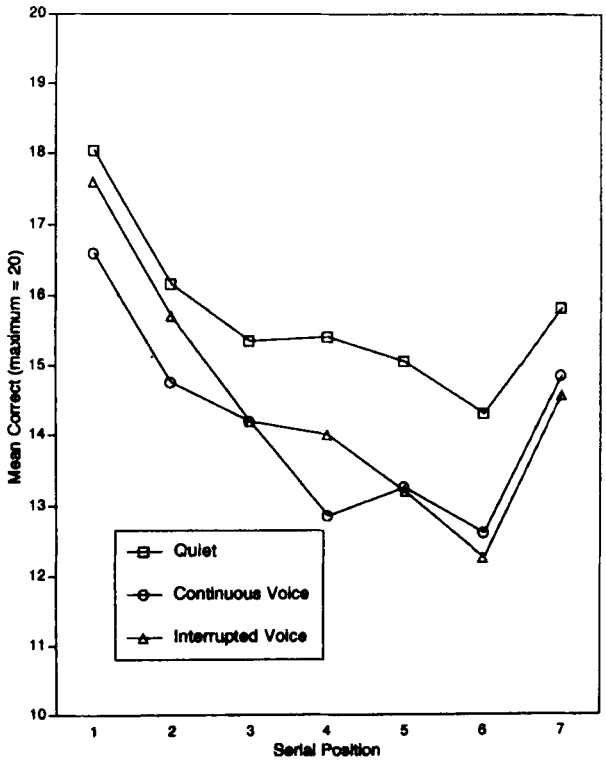

Figure 4. Mean correct responses for each serial position in Experiment 2, contrasting continuously voiced speech with interrupted speech (see Figure 3) and also quiet.

It is quite clear, then, that while the glissandi used in Experiment 1 only disrupted serial recall when they were interrupted by silence, a sequence of varying speech sounds does not require such interruption in order to cause disruption. We propose that streams of (differing) speech sounds are inevitably perceived as being segmented because of cues to transition embedded within and naturally occurring in *narrative* speech, whereas slowly changing nonspeech sounds, which possess no such cues, require the interposition of silence in order to bring about segmentation. We regard speech as being one instance of a class of acoustic stimuli whose members are defined by the fact that they possess certain correlated patterns of pitch and amplitude change: these characteristics are not exclusive to speech, so the class will include both speech and nonspeech stimuli. With the additional assumption of changing state, it becomes possible to account for the patterning of results and to offer an explanation for why not all speech streams disrupt and not all nonspeech streams are benign. In Experiment 3 of this series, we attempted

to demonstrate that perceptual organization of the sound by the auditory system, rather than the passive detection of speech-like characteristics, accounts for the auditory changing-state effect.

EXPERIMENT 3

The usual way to construe the idea of a detector is to think of it as being passive, in the sense that it does not actively process or reconstruct the incoming material (see, e.g., Broadbent, 1958). Certain critical features of the sound are passed, and sounds that do not qualify are excluded. This type of filter or detector would not, by definition, be able to reconstitute a missing part of a signal, for example. Such reconstructive processes would necessarily be different from the notion of filtering as most observers, including Salamé and Baddeley (1989) might interpret it. In Experiment 3, we tested the idea that such a reconstitutive process can be made to influence the degree of disruption of serial recall. This would not be predicted by a passive detector that selects features on the basis of a physical criterion, but only by some mechanism that seeks to organize the incoming information into separate events or objects, as is supposed by the changing state hypothesis.

In Experiment 3, we relied upon a well-established perceptual phenomenon to demonstrate this point. Following the work of Dannenbring (1976; but see also Houtgast, 1972), we used an interrupted pitch glide like that in Experiment 1, but instead of silence, a noise-burst filled the interval between the glides. If the noise has a central frequency and a bandwidth that could potentially mask the glide, subjects should perceive a quite different auditory scene from that produced if the interpolated noise is outside the frequency range of the glide. (Note that the glide and the noise are not contemporaneous, but rather interleaved.) Typically, the noise in the same frequency range as the glide gives rise to the illusion of a continuous glide (accompanied by a now-dissociated sound stream made of regularly occurring noise bursts). A continuous glide is said to be "induced" (see Bregman, 1990; Handel, 1989). In contrast, when the noise is outside the frequency range of the glide, the listener usually perceives a pitch glide *interrupted* by noise.

From the viewpoint of the changing-state hypothesis, perceptual processes that organize the stream would be the crucial feature in object formation. We therefore predicted that when both types of interrupted glide were compared, the phenomenally continuous glide would minimally disrupt serial recall and the phenomenally discontinuous glide would disrupt serial recall in the usual way. The noise should play an *active* role in the perceptual organization of the stream into component objects. This should be a particularly discriminating method, because broad-band noise was used as the medium for causing the illusion of continuity. Categorically, according to the detector model, noise may not pass into phonological memory: Salamé and Baddeley (1989) state that either a filter or a detector system passes material to the phonological store on the basis of its similarity to speech and that "noise does not have the necessary characteristics to pass" (p. 120). Thus, a strong form of Salamé and Baddeley's model would suppose that noise could play no role in determining the effect of the pitch glide on serial recall. However, the changing-state hypothesis predicts that any acoustic feature that can contribute to the segmentation of the stream will result in the disruption of serial recall. It is worth reemphasizing that both types of stimulus are acoustically discontinuous, to the extent that there are sudden transitions from glide to noise, but that the different predictions are about their respective capacities to disrupt serial recall.

In summary, the detector model predicts that since noise is filtered out, all that will reach memory is an interrupted pitch glide; the noise will be wholly irrelevant whatever its bandwidth or center frequency, and both streams will disrupt memory equally. The changing-state model, however, predicts that only the condition in which the pitch glide is phenomenally discontinuous will produce disruption. Also, according to this view, phenomenal continuity will produce effects on serial recall similar to those of actual physical continuity.

Method

Subjects. Twenty graduate and undergraduate students were paid a small honorarium for participating in the experiment.

Apparatus and Materials. Continuous random frequency glides (glissandi) were generated as in Experiment 1.

These basic stimuli were used to provide three types of auditory material. The first type of material, used in the continuous glissandi condition, was provided by the original recordings of the glissandi described above. For the other two types of auditory material, the continuous glissandi were interrupted every 500 msec with 100 msec of silence, using SoundEdit. software. From these interrupted glissandi, two variants were produced. In the first, the silent gaps were filled with a loud burst of noise that had been low-pass filtered at 4 kHz. For the second stimulus type, a burst of noise was high-pass filtered at 4 kHz and used to fill the silent gaps in the interrupted glissandi. The first type of auditory material therefore contained noise that covered the frequency range of the glides themselves and so produced the perception of the glissandi continuing behind a regular sequence of noise bursts (henceforth referred to as the perceptually continuous condition). The second type of material, because it was interrupted by noise that did not cover the frequency range of the glissandi, produced the perception of interrupted glissandi plus a regular sequence of noise bursts (henceforth referred to as the perceptually discontinuous condition).

Design and Procedure. Three auditory conditions were compared: continuous glissandi, perceptually continuous glissandi, and perceptually discontinuous glissandi. The design and procedure were otherwise identical to those of Experiments 1 and 2.

Results

A 3 (auditory condition) × 7 (serial position) ANOVA was carried out on the data. There were significant main effects of both auditory condition [$F(2,38) = 3.79$, $MS_e = 6.92$, $p < .04$] and serial position [$F(6,114) = 9.16$, $MS_e = 7.30$, $p < .001$], but there was no significant interaction [$F(12,228) = 1.14$, $MS_e = 1.70$, $p > .30$].

The serial position curves for the correct responses in each auditory condition are plotted in Figure 5. Planned comparisons revealed significant differences between continuous glissandi and perceptually discontinuous glissandi ($p < .03$) and between perceptually continuous glissandi and perceptually discontinuous glissandi ($p < .03$), but not between continuous glissandi and perceptually continuous glissandi.

The results strongly suggest that perceptual processes actively organize the auditory input, and that the disruption of serial recall is not due to the mere presence of certain signal parameters. The results of Experiment 3 also make an important contribution to our understanding of the perceptual organization of sound. Although many phenomena of auditory streaming have been demonstrated, most of the demonstrations have taken the form of reports by subjects of their perceptual experience. Despite the fact that these data are in other ways adequate, they do not speak to the important point of whether the organization of streams and events occurs at a preattentive level, since, by definition, the phenomena are reported while they are in the "attentional foreground." Thus, it has always been unclear whether the effects of auditory streaming are truly "preattentive." Bregman (1990) views these effects as a result of a preattentive mechanism, as a manifestation of the coexistence of several streams of information, each of which is organized by perceptual rules working outside the focus of attention. Others have regarded auditory stream segregation as a failure of attention (e.g., M. R. Jones, 1976). By showing that auditory induction occurs in a setting in which the subject is explicitly instructed to ignore the sound and is undertaking another task that is attentionally demanding, Experiment 3 provides independent confirmation that such processes occur at a preattentive level. Moreover, the main metric of the auditory induction effect is not in this case a subjective estimate, but an objective measure of the efficiency of serial recall. Therefore, despite Bregman's pessimism that "the issue of whether segregation is preattentive or not will not be able to be resolved until there is some agreement in the research community about what the properties of attention are and how to distinguish experimentally between attention and a process that is preattentive" (Bregman, 1990, p. 209), we believe that the results of Experiment 3 contribute significantly to the conclusion that these processes are preattentive.

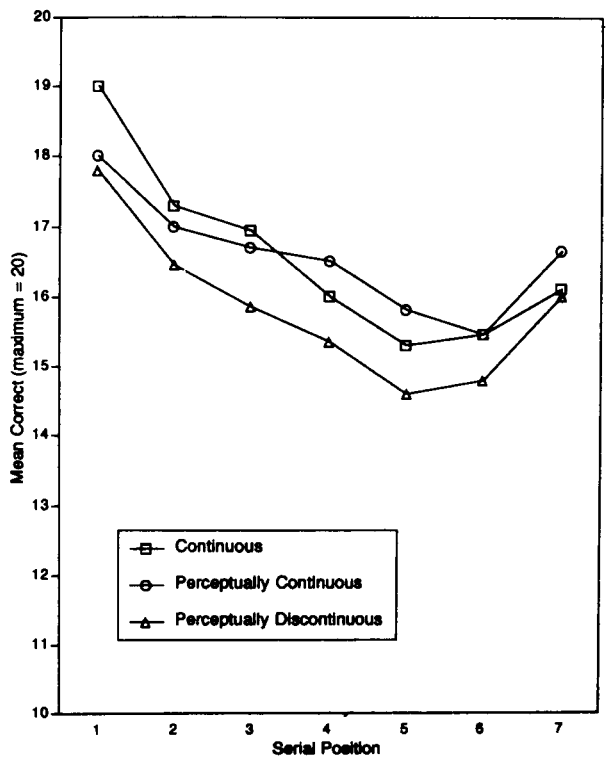

Figure 5. Mean correct responses for each serial position in Experiment 3, contrasting effects of perceptually continuous and perceptually discontinuous glides with physically continuous glides.

GENERAL DISCUSSION

The results of these three experiments point to several major conclusions. First, they reinforce the view that nonspeech stimuli are capable of disruption of serial recall (the "irrelevant speech effect") and extend its compass to include sounds that are not steady state (D. M. Jones & Macken, 1993, have shown that steady-state tones of different frequencies separated by silence show an "irrelevant speech" effect). Second, they show that the notion of changing state does not apply to sounds that are continuous and slowly varying. Third, they show that a nondisruptive, continuous sound can be made to cause disruption by the simple step of inserting periodic intervals of silence that break up the stream. We propose that this occurs with nonspeech because silence serves to segment the stream into a sequence of events, or objects, as opposed to a single temporally extended event. Fourth, they show that silence is not a necessary condition for the same effects to be shown with speech. Fifth, they show that bandpass noise can serve as the basis for segmentation and may then in turn produce disruption of serial recall. Sixth, they show that preattentive processes of perceptual organization (rather than the presence of "critical features" that may be passed by a detector) play a preeminent role in disruption of serial recall.

Segmentation and the Role of Interruption

Whether silence or noise bursts are necessary conditions for segmentation of nonspeech streams remains to be investigated further. Certainly, if we are to portray the basis of the disruption as being the formation of auditory objects, common experience and intuition suggest that these may be discerned even when acoustic stimulation is continuous: the sound arising from a long-continued bowing of a violin, for example, may have within it distinct notes, which surely also have the capacity to achieve the status of an "object." It may have been that the rate

of change of the glides in the temporal domain in Experiment 1 was insufficient to enable such objects to be formed—that is, the rate of variation was simply too slow. We used a cutoff frequency of 0.7 Hz, a value that was intended to correspond to the rate of prosodic variation in speech, but a value of cutoff toward the higher frequencies associated with word or syllable production may have been more appropriate. This possibility could be further investigated by parametric studies in which the cutoff frequency of the low-pass filter is varied. For example, it is possible that as the changes in frequency become more rapid as a result of admitting higher frequencies through the low-pass filter, boundaries will be more sharply defined and hence object formation will be more likely. At higher cutoff frequencies, possibly at the frequency of naturally occurring syllables (4–5 Hz; see Cutler, 1990; Huggins, 1964; Samuel, 1991), such object formation may begin to occur. Whether the crucial factor is the rate of syllable or word production is open to further investigation, but other work showing that attention is preferentially allocated to stressed syllables during speech processing (Pitt & Samuel, 1990) suggests the important role played by segmentation at the syllabic level—that is, at a rate of change in the region of 4–5 Hz.

In unpublished pilot studies, we have already explored the effects of rate of change in the glissandi by manipulating the effect of low-pass frequency and observing the way in which this interacts with the disruption of serial recall, but the results to date are inconclusive. We expected that as the low-pass frequency of the forcing function of the pitch glide was increased to a level above the typical rate of syllables in speech, a continuous pitch glide would form "objects." We reasoned that this should occur because of the sharp transitions that the higher frequency components would allow. The pilot work with subjects listening for the "break-up" of the sound has been encouraging: as the cutoff frequency increases above the 4–6 Hz range, listeners judge that the glides become discontinuous and begin to show object-like properties. In unpublished studies, we have failed to show a clear-cut relationship between increasing low-pass frequency and the degree of disruption of serial recall, but there may be interactions of depth with frequency of modulation that need to be controlled. Another possible strategy would be to derive the forcing function not from a random source but from speech itself. Attempts to show disruption when such signals modulate the *intensity* of a noise carrier have been unsuccessful (Salamé & Baddeley, 1989; D. M. Jones et al., 1992), but the effect of changes in pitch derived in a similar way has yet to be explored.

Music and Irrelevant Speech

The results of Experiment 1 help to reconcile the inconsistent results found with music played during serial recall of visually presented lists. Specifically, Salamé and Baddeley (1989) used a range of vocal and instrumental music stimuli in the irrelevant speech paradigm, but with inconsistent results. On the basis of the results reported here, we may expect that instrumental music should produce the classic "irrelevant speech" effect but that the time course of pitch/amplitude variation will be the important determinant of disruption. In sum, we regard the effect as being dependent on the pattern of variation of the sound rather than on the distinction between speech and nonspeech signals.

Are Speech and Nonspeech Equipotential?

The notion that speech is somehow "special" is a long-standing one (cf. Lieberman, 1984). However, a substantial body of evidence from studies of the perception of sound shows that the distinction between speech and nonspeech is difficult to sustain (see Pisoni & Luce, 1987; Samuel & Tartter, 1986). However, from the point of view of "irrelevant speech effects," empirically the claim is difficult to reject. For example, it is difficult to conceive of circumstances that could completely refute the viewpoint of Salamé and Baddeley (1989). Since they do not specify precise physical conditions that distinguish speech from nonspeech the distinction has to be made ad hoc, on the basis of the disruption of the irrelevant speech paradigm. This absence of an independent criterion of "speech-like" may lead to an unfortunate circularity of explanation, along the following lines: "Stimuli that are speech-like are ones that disrupt serial recall, and stimuli that disrupt serial recall are speech-like." Clearly any arguments that take such a form lack explanatory power. We believe that Experiment 3 provides circumstances for a test of Salamé and Baddeley's (1989) position. Nevertheless, it would be possible to modify the notion of a "speech-like" detector still further, so that stimuli that would not themselves alone qualify for entry via the detector qualify for entry when in combination. However, in our view it does not seem to be possible to specify such combinations a priori.

We believe that the argument for according special status to speech within memory cannot be sustained. Instead, we propose that streams that disrupt serial recall do so because they are segmented into events or objects and that each object is different from the one preceding. If this proposal is to be convincing, we must explain how it is that the set of disruptive signals includes speech and sung words, neither of whose acoustic signals is always segmented by *silence* in any obvious way. Clearly we must suggest some mechanism that results in a perceptual process that treats these streams differently from other continuous acoustic signals. There are two possible standpoints from which such an explanation might be attempted; they differ from each other in the level at which the perceptual mechanism might operate. The first rests on the argument that speech streams may be segmented and categorized via high-level recognition processes governed by learned linguistic-phonetic constraints (indicating that speech streams are "special cases"). The second approach suggests that correlated changes in par-

ticular parameters of the signal act as an effective means of segmentation at a more peripheral level (suggesting that any signal will potentially be subject to these processes). According to this "correlated attribute" argument, the multidimensional aspect of speech, and particularly the fact that changes in frequency and amplitude are correlated in the temporal domain, sets it (and any other signals that have these characteristics) apart from unidimensional stimuli in which silence may be the most potent method of segmentation. Correlated changes in frequency and amplitude over time provide a coherence of form that may readily be transformed into a sequence of objects, whether or not the signal is interrupted by silences.

Evidence from studies of speech perception also suggests that the discrimination among classes of speech sounds is based on a mechanism similar to the one we have been advocating for object formation. For example, in the discrimination of stops and glides, Walsh and Diehl (1991) found that both rise time and duration of the formant transition act synergistically. Moreover, the cues used in distinguishing stops from glides in speech are similar to those used in making discriminations among nonspeech stimuli. They conclude that their results "are not likely to be explained strictly in terms of perceivers' previous learning of the *correlated attributes* of stops and glides. If the *natural correlation* between transition duration and rise time enhances the contrast between stops and glides, it is apparently because of acoustic/auditory factors not unique to the perception of speech" (Walsh & Diehl, 1991, p. 614, emphasis added).

There is some evidence that suggests that the other possibility, that of higher level phonetic *categorization* processes of speech perception based on recognition, plays a relatively minor role in segmentation of the speech stream as far as the auditory changing-state effect is concerned. Reversed narrative speech, which does not contain familiar speech segments, produces disruption identical to that produced by narrative English (D. M. Jones, Miles, & Page, 1990). If the special effects found with speech were the result of segmentation into familiar categorical units, we would predict that reversed speech should cause less disruption of recall performance. On the other hand, reversed speech does contain signals that are readily segmented on the basis of the correlated frequency amplitude and time course alone.

Is the Effect One of Distraction?

Any idea that the effects found here are those that we may loosely refer to as "distraction" can be ruled out by three converging lines of evidence that show that the effect of irrelevant speech is on the contents of memory rather than on the process of encoding the visually presented material. First, if the presentation of speech is confined to either the time when the items are presented or the time when they are being rehearsed, then roughly equivalent degrees of disruption are produced. If the effect occurs on encoding alone, it should be confined to cases in which exposure has occurred during the period of presentation (Miles et al., 1991). Second, only memory tasks that require *serial* recall show effects of irrelevant speech. This suggests that the effect depends on the operations undertaken in memory, not on the conditions of presentation. If, for example, the subject is given a probe recall task in which the item following the probe is to be reported (for which serial order information is required), irrelevant speech is disruptive. If no test of *serial* order information is involved, no irrelevant speech effect occurs. So, for example, if the subject is given a list from a fixed set (such as days of the week) from which an item is missing and is then required to report the missing item, there is no effect of irrelevant speech (D. M. Jones & Macken, 1993; Morris & D. M. Jones, 1992). Similarly, Salamé and Baddeley (1990) found no irrelevant speech effects with free recall. Also, Morris and D. M. Jones (1990a) demonstrated that memory for the position of dots was not disrupted by irrelevant speech, a finding that can be accommodated by supposing that spatial material is stored on a visuospatial scratch pad rather than in phonological working memory. Third, the attention-demanding Stroop color-naming task is not usually susceptible to the presence of irrelevant speech (Miles & D. M. Jones, 1989; Miles, Madden, & D. M. Jones, 1989; Thackray & K. N. Jones, 1971; Thackray, K. N. Jones, & Touchstone, 1972), a fact that also suggests that the effect is not one at the encoding of visual material. Additionally, we have a number of unpublished studies showing no effects of irrelevant speech on tasks that do not involve storing materials in memory, including simple nonverbal tasks requiring sustained attention (see Jones, in press, for a review).

The Object-Oriented Episodic Record Model

The notion of segmentation of the auditory stream and object formation fits in well with our general theoretical perspective that these and other working memory phenomena can be explained in terms of a blackboard analogy. Several other approaches also rely on the notion of a virtual surface on which objects are represented. Among the most powerfully articulated are Anderson's (1983) ACT model, Kahneman and Treisman's (1984) "object file," and Marr's (1976) "object tokens." Generally, these theories have been preoccupied with object formation in the visual rather than the auditory domain, and with isolated events rather than ones for which serial order must be retained.

Like other formulations of short-term storage, we envisage the blackboard in the O-OER model as an area for temporary storage in which there is relatively rapid decay. However, we depart from convention by supposing that the *links between* objects are subject to decay, rather than the objects themselves. Essentially, retrieval consists of reproducing an episodic record that comprises objects—abstract representations of items, undifferentiated by their modality of origin, for example—and these objects are linked sequentially by pointers. When lists of to-be-remembered items are presented visually, covert ar-

ticulation establishes these links. Background auditory stimulation also produces objects on the blackboard through detection of changing state. The process of retrieval is undertaken by a production system that uses the pointers to trace the episodic record, or, to use Rumelhart's (1991) concept, "episodic trajectory." We see the process of changing state as setting down competing trajectories to the one laid down by rehearsal of the to-be-remembered list. Since items are undifferentiated according to their modality of origin, these linkages are the only means by which retrieval may be undertaken (see D. M. Jones, in press, for a general discussion). Novel predictions may be made on the basis of this "equipotentiality" hypothesis. Among them is the prediction that articulatory suppression will show changing-state effects, since according to the hypothesis the representation of subvocal speech is the same as that for speech from an external source. Yet another possibility, representing a strong form of the hypothesis, is that memory for nonverbal visual events, provided that a test of their serial order is required, can be disrupted by irrelevant speech (hitherto, tests of the irrelevant speech effect have consisted of spatial tasks in which memory for order of visual events was not tested; see Morris and Jones, 1992).

The nature of production rules within the production system may be usefully addressed through reference to the psychophysics of auditory event perception (see, e.g., Bregman, 1990; Handel, 1989; Warren, 1982). This work has shown how certain types of continuity in the acoustic stimulus may lead to the perception of a single continuous auditory event, whereas certain types of discontinuity serve to segment the acoustic stimulus into separate auditory events. So, within the present formulation, acoustic stimuli that are perceived as representing a single continuous event will be registered in short-term memory as a single coherent auditory object (since such a stream constitutes only a single object, it does not require any pointers to maintain links *between* objects) that will be less likely to interfere with information of visual origin, whereas acoustic stimuli that are likely to be segmented by the perceptual process will lead to the registration of discrete objects within short-term memory that will be connected by rather weaker or more ambiguous serial order pointers, and will be more likely to cause disruption of visual short-term recall (this is what was shown in Experiment 1 of the series). Moreover, *boundaries* (such as end of list or end of rehearsal group) are particularly well signified by pointers (see Frankish, 1985; Penney, 1989; Seamon & Chumbley, 1977, for discussion of serial recall processes).

Through examination of some of the acoustic parameters that determine the disruption of serial recall, in the present experiments we have attempted to illuminate attentional processes in short-term recall. These experiments have strongly reinforced the view that nonspeech materials produce disruption that was previously believed to be confined to speech stimuli. The experiments have also pointed to the preattentive nature of auditory streaming. Additionally, the findings, together with convergent evidence from studies of speech perception have helped to clarify the status of the speech/nonspeech distinction. Much remains to be done to flesh out the details of the O-OER model, possibly by focusing on task parameters and by examining other interfering factors such as articulatory suppression.

REFERENCES

ANDERSON, J. R. (1983). *The architecture of cognition*. Cambridge, MA: Harvard University Press.

BREGMAN, A. S. (1990). *Auditory scene analysis*. Cambridge, MA: MIT Press.

BROADBENT, D. E. (1958). *Perception and communication*. Oxford: Pergamon.

CUTLER, A. (1990). Exploiting prosodic probabilities in speech segmentation. In A. T. M. Altmann (Ed.), *Cognitive models of speech processing: Psycholinguistic and computational perspectives* (pp. 105-201). Cambridge, MA: MIT Press.

DANNENBRING, G. L. (1976). Perceived auditory continuity with alternately rising and falling frequency transitions. *Canadian Journal of Psychology*, **30**, 99-114.

FRANKISH, C. R. (1985). Modality-specific grouping effects in short-term memory. *Journal of Memory & Language*, **24**, 200-209.

HANDEL, S. (1989). *Listening*. Cambridge, MA: MIT Press.

HOLMES, J. (1985). A parallel-formant synthesiser for machine voice output. In F. Fallside & W. A. Woods (Eds.), *Computer speech processing* (pp. 163-187). London: Prentice-Hall.

HOLMES, J. (1986). *Syncon: A synthesis by rule software package for convenient interactive control of the LSI speech synthesiser using a BBC microcomputer* [Computer program]. Available from J. Holmes, 19 Maylands Drive, Middlesex UB8 1BH, England.

HOLMES, J., MATTINGLY, I. G., & SHEARME, J. N. (1964). Speech synthesis by rule. *Language & Speech*, **7**, 127-143.

HOUTGAST, T. (1972). Psychophysical evidence for lateral inhibition in hearing. *Journal of the Acoustical Society of America*, **51**, 1885-1894.

HUGGINS, A. W. F. (1964). Distortion of the temporal pattern of speech: Interruption and alternation. *Journal of the Acoustical Society of America*, **36**, 1055-1064.

JONES, D. M. (in press). Objects, streams and threads of auditory attention. In A. D. Baddeley & L. Weiskrantz (Eds.), *Attention: Selection, awareness and control*. Oxford: Oxford University Press.

JONES, D. M., & BROADBENT, D. E. (1991). Human performance and noise. In C. M. Harris (Ed.), *Handbook of acoustical measurements and noise control* (pp. 24.1-24.24). New York: McGraw-Hill.

JONES, D. M., & MACKEN, W. J. (1993). Irrelevant tones produce an "irrelevant speech effect": Implications for phonological coding in working memory. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **19**, 1-13.

JONES, D. M., MADDEN, C., & MILES, C. (1992). Privileged access by irrelevant speech to short-term memory: The role of changing state. *Quarterly Journal of Experimental Psychology*, **44A**, 645-669.

JONES, D. M., MILES, C., & PAGE, J. (1990). Disruption of proof-reading by irrelevant speech: Effects of attention, arousal or memory? *Applied Cognitive Psychology*, **4**, 89-108.

JONES, D. M., & MORRIS, N. (1992). Irrelevant speech and cognition. In D. M. Jones & A. P. Smith (Eds.), *Handbook of human performance* (pp. 29-53). London: Academic Press.

JONES, D. M., & MORRIS, N. (in press). Irrelevant speech and serial recall: Implications for theories of attention and working memory. *Scandinavian Journal of Psychology*.

JONES, M. R. (1976). Time, our lost dimension: Toward a new theory of perception, attention and memory. *Psychological Review*, **83**, 323-355.

KAHNEMAN, D., & TREISMAN, A. (1984). Changing views of attention and automaticity. In R. Parasuraman & D. R. Davies (Eds.), *Varieties of attention* (pp. 29-61). London: Academic Press.

KELLER, E. (1990). *Signalyze: Signal analysis for speech and music*. Seattle: InfoSignal Inc.

LIEBERMAN, P. (1984). *The biology and evolution of language*. Cambridge, MA: Harvard University Press.

MARR, D. (1976). Early processing of visual information. *Philosophical Transactions of the Royal Society of London: Series B*, **275**, 483-524.

MILES, C., & JONES, D. M. (1989). The fallacy of the cross-modal Stroop effect: A rejoinder to Cowan. *Perception & Psychophysics*, **45**, 82-84.

MILES, C., JONES, D. M., & MADDEN, C. (1991). Locus of the irrelevant speech effect in short-term memory. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **17**, 578-584.

MILES, C., MADDEN, C., & JONES, D. M. (1989). Cross-modal, auditory-visual Stroop interference: A reply to Cowan and Barron. *Perception & Psychophysics*, **45**, 77-81.

MORRIS, N., & JONES, D. M. (1990a). Habituation to irrelevant speech: Effects on a visual short-term memory task. *Perception & Psychophysics*, **47**, 291-297.

MORRIS, N., & JONES, D. M. (1990b). Memory updating and working memory: The role of the central executive. *British Journal of Psychology*, **81**, 111-121.

MORRIS, N., & JONES, D. M. (1992). *Multiple resources in verbal short-term memory*. Manuscript submitted for publication.

MORRIS, N., QUAYLE, A., & JONES, D. M. (1987). Memory disruption by background speech and singing. In E. Megaw (Ed.), *Contemporary ergonomics* (pp. 494-499). London: Taylor and Francis.

PENNEY, C. G. (1989). Modality effects in delayed free recall and recognition: Visual is better than auditory. *Quarterly Journal of Experimental Psychology*, **41A**, 455-470.

PISONI, D. B., & LUCE, P. A. (1987). Acoustic-phonetic representations in word recognition. *Cognition*, **25**, 21-52.

PITT, M. A., & SAMUEL, A. G. (1990). Attentional allocation during speech perception: How fine is the focus? *Journal of Memory & Language*, **29**, 611-632.

RUMELHART, D. (1991). *Connectionist concepts of learning, memory and generalization*. Paper presented at the International Conference on Memory, Lancaster University.

SALAMÉ, P., & BADDELEY, A. D. (1982). Disruption of short-term memory by unattended speech: Implications for the structure of working memory. *Journal of Verbal Learning & Verbal Behavior*, **21**, 150-164.

SALAMÉ, P., & BADDELEY, A. D. (1989). Effects of background music on phonological short-term memory. *Quarterly Journal of Experimental Psychology*, **41A**, 107-122.

SALAMÉ, P., & BADDELEY, A. D. (1990). The effects of irrelevant speech on immediate free recall. *Bulletin of the Psychonomic Society*, **28**, 540-542.

SAMUEL, A. G. (1991). Perceptual degradation due to signal alternation: Implications for auditory pattern processing. *Journal of Experimental Psychology: Human Perception & Performance*, **17**, 392-403.

SAMUEL, A. G., & TARTTER, V. C. (1986). Acoustic-phonetic issues in speech perception. *Annual Review of Anthropology*, **15**, 247-273.

SEAMON, J. G., & CHUMBLEY, J. I. (1977). Retrieval processes for serial order information. *Memory & Cognition*, **5**, 709-715.

THACKRAY, R. I., & JONES, K. N. (1971). Level of arousal during Stroop performance: Effects of speed stress and "distraction." *Psychonomic Science*, **23**, 133-135.

THACKRAY, R. I., JONES, K. N., & TOUCHSTONE, R. M. (1972). The color-word interference test and its relation to performance impairment under auditory distraction. *Psychonomic Science*, **28**, 225-227.

WALSH, M. A., & DIEHL, R. L. (1991). Formant transition duration and amplitude rise time as cues to the stop/glide distinction. *Quarterly Journal of Experimental Psychology*, **43A**, 603-620.

WARREN, R. M. (1982). *Auditory perception: A new synthesis*. New York: Pergamon.

(Manuscript received November 1, 1991;
revision accepted for publication August 27, 1992.)