Organizational Factors in Selective Attention: The Interplay of Acoustic Distinctiveness and Auditory Streaming in the Irrelevant Sound Effect

Dylan Jones and David Alford
Cardiff University

Sébastien Tremblay and Bill Macken
Cardiff University

Andrew Bridges
University of Aston

A series of studies further explored the way in which irrelevant sound disrupts the serial recall of visually presented verbal sequences. The hypothesis that distinctiveness (stimulus mismatch) within auditory irrelevant sequences is a critical determinant of disruption of serial recall was tested. Experiment 1 showed that the degree of disruption was related to the degree of mismatch between successive stimuli. However, in Experiment 2, changes in 2 attributes of a stimulus produced less disruption than when only 1 was changed, suggesting mismatch alone was not the key factor. These results were reconciled with the changing-state hypothesis in Experiment 3 in which change and disruption were monotonically related up to the point at which mismatch created 2 streams. Object-based theories are able to explain this pattern of results.

Short-term serial recall for visually presented sequences is impaired markedly by the presence of sound irrelevant to the task, even though participants are asked to ignore the sound (e.g., Buchner, Irmen, & Erdfeelder, 1996; Colle & Welsh, 1976; Hellbrück, Kuwano, & Namba, 1996; Salamé & Baddeley, 1982; see Ellermeier & Zimmer, 1997, for a useful discussion of the magnitude of the effect). It has been suggested that disruption is the product of a clash between two concurrent processes of seriation, one from the deliberate rehearsal of items within the serial recall task, the other a by-product of automatic and preattentive organization of irrelevant sound into streams (Jones, 1994; Jones, Beamam, & Macken, 1996). Interference arises not because of a similarity in the identity of the items in memory to those in the irrelevant sequence; rather, it occurs because the serial recall task and the processing of sound both involve processes of seriation (Jones & Macken, 1995b; Macken, Mosdell, & Jones, in press). The functional character of the memory task is not the subject of the present article (but see Beaman & Jones, 1997, 1998); instead, the character of the irrelevant stream, particularly the role played by the physical mismatch between successive stimuli and the relation of this mismatch to seriation, is highlighted. Specifically, interest centers on the hypothesis that the degree of mismatch between successive stimuli (the distinctiveness of stimuli in the sequence) is related to the degree of disruption of serial recall.

The phenomena associated with the irrelevant sound paradigm assume importance for the study of both attention and memory. Given that participants are instructed to ignore the irrelevant sound, and hence arguably the material is therefore unattended, the extent of interference with memory may reveal the nature and extent of preattentive processing of sound. That there is interference under these conditions implies that the processing of sound is obligatory and that there is some shared resource or process being addressed. In addition to having a number of implications both for theoretical issues in selective attention (classically, addressed through variants of the dichotic listening paradigm) the form of this obligatory processing will have implications for a range of practical issues such as noise abatement and the design of auditory alarms. The work also has a range of implications for auditory perception. For example, the available evidence indicates that those organizational factors found when attention is focused deliberately are also manifest in this obligatory processing (see Jones et al., 1996). More generally, the work has implications for the global representation of cognitive architecture. If the irrelevant sound is processed without conscious awareness at a precategorical level, then the mere fact that it interferes with deliberate rehearsal in memory (Beamam & Jones, 1997, 1998) suggests that precategorical analysis and postcategor-
tical levels of processing are less distinct than is commonly assumed (see, e.g., Pashler, 1998). Broadly, research on the effects of irrelevant sound has highlighted the intimacy of the connection between perception and short-term memory phenomena (see Cowan, 1995, for a related view). An emerging generalization is that the process of ordering of events at the preattentive level in the perception of sound interferes with the order of events rehearsed in memory. This article seeks to specify this process more fully.

There is relatively strong evidence, converging from several sources, against the idea that disruption of serial recall is based on the similarity of content between the to-be-recalled material and the irrelevant sound. For example, given that the material rehearsed in memory is almost invariably verbal, one might expect, therefore, that disruption would be confined to irrelevant sequences containing verbal material. But it has been shown that speech is neither a necessary condition because some speech sounds do not produce disruption (Jones & Macken, 1993; Jones, Macken, & Murray, 1993; Jones, Madden, & Miles, 1992; LeCompte, Neely, & Wilson, 1997), nor a sufficient condition, because music and tones also produce disruption (Jones & Macken, 1995b; Jones et al., 1993; Klatte, Kilcher, & Hellbrück, 1995; Salamé & Baddeley, 1989). When the similarity of the content of the irrelevant sound to the content of the to-be-remembered sequence is manipulated directly, the degree of disruption does not vary; rather, disruption increases as the degree of similarity within the irrelevant sequence decreases (Jones & Macken, 1995c; Martin-Loeches, Schweinberger, & Sommer, 1997). For example, rhyming sequences (such as key, flea, tea) are less disruptive than nonrhyming sequences (deaf, pay, bell), regardless of whether they also rhyme with items in the to-be-remembered list. The fact that nonrhyming sequences are more damaging to serial recall suggests that irrelevant sequences made up of more distinct elements are more disruptive (Jones & Macken, 1995c). Conventionally, within the domain of memory, ease of retrieval is associated with interference of representation at the item level (embodied in concepts such as pro- and retroactive interference). The research on irrelevant speech suggests that process (particularly the organization of items into a series) is at least as important as content.

Other lines of evidence point to the importance of change within the irrelevant sequence as the key determinant of disruption. For example, a repeated sequence of sounds produces much less disruption relative to quiet than does a sequence containing marked changes (Jones et al., 1992; Jones & Macken, 1993, 1995a, 1995b, 1995c; LeCompte, 1994, 1995; Tremblay & Jones, 1998, in press). These findings, along with the studies of rhyming already mentioned, contributed to the formation of the changing-state hypothesis. Essentially the hypothesis predicts that any auditory stimulus, whether speech or nonspeech, as long as it shows some appreciable change from two immediately successive tokens (or stimulus mismatch), will produce disruption of serial recall (Jones et al., 1992; Tremblay & Jones, 1998, in press). Conceptually, the notion of mismatch is mapped onto the representation of state. It is argued that, if there is no mismatch, only one representation of the (repeated) event is established within memory; generally, if successive events exhibit mismatch, different representations of each event are produced and these are coupled by linkages (or vectors) representing their order. A strong form of the changing-state hypothesis suggests that the greater the mismatch, the less ambiguous the linkage, the more that changing state is represented, and in turn the more marked is the disruption of serial recall.

In the experiments that follow, we attempted to show how the degree of distinctiveness of successive items in the auditory stream is related to the disruption of serial recall.

**Experiment 1**

In this experiment, we undertook a simple test of the proposition that sequences that are more distinct acoustically are more disruptive than those that are less distinct; we did this by manipulating the degree of variation in frequency of a sequence of pure tones. These simple stimuli were chosen on the grounds that a single salient feature (namely, frequency) could be varied in isolation. Three types of sequences containing pure (sine) tones were compared: one in which there was a relatively small difference between each event in the sequence, another in which there were slightly larger differences, and a third in which there were large differences between the tones. The prediction, based on the changing-state hypothesis, was that the differences become larger so should the degree of disruption of serial recall.

**Method**

**Participants.** Twenty-five male and female undergraduate student volunteers at the University of Aston took part in the experiment as part of a course requirement. All participants had English as their first language, had normal or corrected-to-normal vision, and reported normal hearing.

**Apparatus and materials.** Lists of items to be recalled were presented serially on the screen of an Apple Macintosh Performa 6200 microcomputer. Lists were constructed from the random arrangement of the letters f, k, l, m, q, r, s, t, and y, with the constraints that a letter could not appear in the same serial position in two consecutive lists and that recognizable words, acronyms, or letter strings were excluded. The lists were stored and presented within a SuperCard (Version 2.5, Allegiant, San Diego, CA) environment.

The auditory stimuli were constructed from three sets of nine tones, produced and edited using SoundEdit Pro software (Version 1.0, Macromind and Paracomp, San Francisco, CA). Three types of irrelevant sound sequences were assembled, on the basis of the frequency difference between members of each set, namely small, medium, and large. In the small-difference condition, the tones had frequencies of 171, 174, 177, and 180 Hz; in the medium-difference condition, the tones had frequencies of 150, 171, 315, and 673 Hz; and in the large-difference condition, the tones had frequencies of 87, 174, 348, and 696 Hz. In the latter case, each tone was an octave apart.

All tones were edited using digital signal processing techniques to last exactly 250 ms, with a rise time of 10 ms, and separated by 250 ms of silence. The number of tones was the same in each condition (see Bridges & Jones, 1996). The ordering of the tones
was random, with the constraint that each member of the set had to be used before any member of the set could be repeated. Within these limits, a recording was constructed for each condition lasting approximately 20 s. Sound was delivered via headphones at 65 dB(A) as measured by an artificial ear (Bruel & Kjaer, Naerum, Denmark).

Procedure. A repeated measures design was used in the experiment, with all participants undertaking the recall task under all four acoustic conditions (large, medium, and small difference, coupled to a quiet control). Participants were tested individually, seated in a soundproof laboratory approximately 0.5 m from the computer screen. At the outset, standard instructions were read by the participant. These informed them of the nature of the recall task and asked them to ignore any sounds they might hear. In each trial, the nine letters were displayed in random order as described above. Each letter was displayed for 0.5 s with an interstimulus interval also of 0.5 s. Every list was preceded by a warning tone. After presentation of the last item in the list, the word wait was displayed on the screen for 10 s during which the participant was expected to rehearse the list. After this retention interval, once the word wait disappeared, the participant was required to recall the list in strict serial order. Sound was played throughout both presentation and retention of the lists. Responses were written on a blank grid composed of rows of nine boxes each. The participants were given 15 s to recall each list, after which the next trial would start automatically. Sixteen lists were presented in each condition, making a total of 64 lists. The lists were split up into four blocks of 16 lists. The ordering of the conditions was randomized from trial to trial within each block, with the constraint that the condition for each trial differed from the one that preceded it. The experimental trials were preceded by a short practice session, and the experiment lasted approximately 45 min.

Results and Discussion

Responses were scored in terms of a strict serial recall criterion: The correct item had to be in the correct position for it to be scored as correct. The overall mean errors of recall in each condition totaled 7.36 when the task was performed in quiet, 7.66 in the small-difference condition, 7.90 in the medium-difference condition, and 8.44 in the large-difference condition. The results are in the order predicted by the changing state hypothesis, with a gradual increase in error as the difference between the tones increased. Errors were analyzed with a two-variable repeated measures analysis of variance (ANOVA), with acoustic condition (four levels: quiet, and small, medium, and large differences) and serial position (nine levels) as variables. There were significant main effects of acoustic condition, \( F(3, 72) = 5.93, \text{MSE} = 47.66, p = .001 \), and serial position, \( F(8, 192) = 39.93, \text{MSE} = 529.25, p < .0001 \), with no significant interaction between the two variables (\( F < 1 \)). The means at each serial position are displayed in Figure 1 (here, as elsewhere in the series, error bars based on standard errors are given).

Planned comparisons showed significant differences between quiet and two of the sound conditions: quiet versus medium, \( F(1, 24) = 4.18, p < .05 \), and quiet versus large, \( F(1, 24) = 16.59, p < .0001 \). The difference between quiet and small was not significant (\( p = .26 \)). Performance in the large condition was significantly worse than that in the other two sound conditions; hence, there were significant effects of the contrast of large versus small conditions, \( F(1, 24) = 8.56, p < .01 \), and of medium versus large conditions, \( F(1, 24) = 4.11, p < .05 \). There were no other significant differences between auditory conditions.

Clearly, the results indicate increased disruption across the auditory conditions. Most disruption was produced by the condition containing tones with the greatest frequency difference from one event to the next; in the condition in which the frequency differences were small, performance did not differ significantly from that in quiet.

The results of Experiment 1 are encouraging for the changing-state hypothesis and are consistent with the results found with rhyming and nonrhyming stimuli (Jones & Macken, 1995c).

Experiment 2

In Experiment 2, we tested a further embodiment of the changing-state hypothesis, that because the process of mismatch is supposed to be based on physical (spectral) cues, variation in the acoustic parameters of a sound can act additively. This is a not unreasonable assumption given that acoustic cues can act additively in auditory perception (cf. Bregman, 1990). Therefore, if the two attributes of the stimulus vary, the degree of mismatch between successive stimuli becomes greater than if one parameter varies. This should mean that the disruption of serial recall becomes more marked when two are varied than when one is varied. In the present study, tones exhibiting changes either in timbre (harmonic content or spectrum) only, in pitch (fundamental frequency) only, or in both timbre and pitch are used as irrelevant stimuli. Specifically, we varied the complexity.
of the timbre by using sounds that were either sinusoids (which have acoustic energy only in one component, the fundamental, and have no harmonics), square waves (which have energy only at odd harmonics), or sawtooth waves (which have energy at all harmonics). Physically, these three types of stimuli represent points along a continuum of increasing harmonic complexity. Phenomenally, the three types of waveform are distinct and the continuum reflects an increasing loss of "purity" of the sound, with the sine wave producing greatest clarity, the square wave producing a rougher sound, and the sawtooth producing a more jagged sound.

The three types of signals, when placed randomly within a sequence, represent a mismatch in terms of the timbral quality of the sound (or, in acoustic terms, the contrast of harmonic complexity). When each of the sounds is played at the same fundamental frequency, the frequencies of each of the harmonics will be the same for each sound (although the distribution of energy across those harmonics will be different). However, if we covary fundamental frequency with timbre (or harmonic composition), the contrast between successive stimuli will be more marked because the spacing as well as the number of harmonics will change. We expected the joint variation of pitch and timbre to act additively, therefore, and to produce marked stimulus mismatch; hence, their joint action should be particularly damaging for serial recall.

Method

Participants. Twenty-six students took part in this experiment in exchange for a small honorarium. All reported normal hearing and normal (or adjusted-to-normal) vision.

Apparatus and materials. The memory task was identical to the task used in Experiment 1.

Irrelevant sound sequences were made up of sequences of sounds of one or more of three frequencies (220, 440, and 880 Hz) or three types of signals (sine, square, and sawtooth waves). A set of nine stimuli (three types of signals at each of three different frequencies) allowed the creation of four different sequences corresponding to the four auditory conditions: (a) repeated frequency, repeated timbre—only one of the tones was repeated (both frequency and timbre held constant within a sequence); (b) repeated frequency, changing timbre—composed of a series of three tones exhibiting changes in timbre only (all three were 440-Hz tones) in a random order; (c) changing frequency, repeated timbre—a series of three tones presented in random order within a triplet with frequencies at octave intervals of 220 Hz, 440 Hz, and 880 Hz (timbre held fixed within a sequence); and (d) changing frequency, changing timbre—both frequency and timbre changed randomly. All changing sequences were edited with the constraint that two successive tokens were not the same.

Experimental design. There were five conditions in all: the four auditory conditions just described and a quiet control. A repeated measures design was used in which presentation of trials was prearranged quasi-randomly such that the five conditions varied over successive trials. There were 15 trials per condition, 75 trials in all.

Procedure. The procedure was generally similar to the previous experiment in the current series. The experiment took 50 min.

Results and Discussion

The overall form of the results (except for the control condition in which the mean error was 36%) is depicted in Figure 2. In the pair of columns on the left is the usual result, one for which there was precedent in Experiment 1 of the current series as well as elsewhere: When timbre was held constant, changing its pitch increased the error appreciably. However, in the pair of columns on the right the results show that the effect was not as predicted by a simple form of the changing-state hypothesis: Repeating pitch, changing timbre produced quite marked disruption; but when both attributes were changing, the level of disruption diminished. Having more changes in the physical character of the sound reduced disruption, a result at variance with the prediction of the changing-state hypothesis.

Statistical analysis of the serial recall errors supported this portrayal of the results: A 5 (auditory condition) × 9 (serial position) repeated measures ANOVA was undertaken in which the main effects of acoustic conditions were found to be significant, $F(4, 100) = 6.23, MSE = 2.81, p < .001$, as was serial position, $F(8, 200) = 33.83, MSE = 7.07, p < .0001$; but the interaction of acoustic conditions and serial position was not significant ($F < 1$).

The four critical auditory conditions were also contrasted in a 2 (repeated/changing pitch) × 2 (repeated/changing timbre) ANOVA with the quiet control removed. Main effects were not significant: neither for pitch, $F(1, 25) = 0.41, MSE = 1.25, p = .52$; nor timbre, $F(1, 25) = 0.07, MSE = 0.44, p = .70$; but the interaction of pitch and timbre was significant, $F(1, 25) = 8.37, p < .01$. The composition of the interaction was revealed by simple main effects (Dunn’s $\alpha$ of .04 was calculated and used in order to compensate for the familywise error) showing a significant

![Figure 2](image-url)
disruptive effect of changing frequency at a constant timbre, $F(1, 25) = 5.17, M SE = 1.30, p < .03$; but it did not reach significance when timbre was also changing, $F(1, 25) = 2.49, M SE = 0.99, p = .13$. Recall performance was significantly impaired by changing timbre when frequency was repeated, $F(1, 25) = 7.12, M SE = 0.50, p < .01$; but when frequency was changing as well, the significant effect was one of facilitation relative to the effect of changes in timbre alone, $F(1, 25) = 5.28, M SE = 0.97, p < .05$.

Varying tones in two dimensions produced much less disruption on serial recall than varying tones in a single dimension, in the form of changes in either pitch or timbre. Clearly, the results of Experiment 2 are at variance with the results of Experiment 1. Perhaps the key to understanding these contrary effects lies in an understanding of the modulating influence of auditory stream formation and its consequences for seriation. One generalization that may make intelligible the results of the foregoing series is that the representation of order changes at different degrees of mismatch. That is, the degree of mismatch will yield order information only as the degree of mismatch is increased up to a certain point; beyond that, order information will diminish. In abstract terms, if the mismatch retains the essential "object" character of the sequence, then order information within a sequence is preserved; when the difference between events is so large as to invoke a series of sounds, each potentially from a different object, the order information is diminished.

Greater weight may be placed on this analysis because it fits closely with an already established literature on the perception of order in sequences of sound. A seminal finding in this research area is that untrained listeners cannot report the order of recycled sequences of four unrelated sounds—for example, a tone, a burst of white noise, a vowel sound, and a buzz—even though the sounds were clearly discriminable one from another (Warren & Obusek, 1972; Warren, Obusek, Farmer, & Warren, 1969; see also Broadbent & Ladefoed, 1959). However, if the sequences are modified so that there are repetitions within the sequence—so that it comprises a tone of low pitch, a tone of high pitch, a white noise burst, and a buzz—there is a dramatic increase in the listeners' capacity to detect order. In sequences in which high and low tones were contiguous, the identification of order became quite good (68% correct), but when the tones were not contiguous, identification was rather poor (20% correct). It is argued that this dramatic improvement was brought about by the fact that differences in order can be detected in contiguous stimuli that are different in one feature but not so different as to belong to a different timbral class (see also McNally & Handel, 1977). In these circumstances, knowing the order of two tones in the sequence also allows the disambiguation of the order of others in the sequence.

The studies of memory for order in sequences of sounds carry with them two implications for our interest in the effects of irrelevant sounds. First, they point to the fact that empirically we might expect that marked differences in stimulus mismatch will have quite different consequences for the representation of order of the irrelevant stream. Second, given the emphasis within the changing-state hypothesis on the process of seriation in the irrelevant sequence, they suggest that the empirical pattern of interference should produce a nonmonotonic relation between mismatch and the degree of seriation, and hence the degree of disruption of serial recall.

As the members of a sequence share fewer and fewer characteristics, there is an increasing tendency for each of the elements not to be perceived as part of coherent whole. This may be illustrated by the way in which streaming by pitch occurs. Consider a long sequence of sound in which there are stimuli that alternate in pitch and in which the difference in pitch is gradually increased. When the pitch differences are rather small, listeners tend to perceive a sound sequence undulating in pitch; as the difference in pitch is increased further, a galloping rhythm is perceived, even though the objective rate of presentation remains fixed. At each of these levels of pitch difference, the percept remains coherent; the variations in pitch are usually perceived as belonging to a single perceptual entity. However, as the difference becomes more marked, perceptual fission occurs. At high levels of pitch difference, two temporally extended entities are formed, one of high pitch and one of low pitch, each unvarying in pitch, each regarded by the listener as a different perceptual object. Such effects have been known and exploited for some considerable time (notably in the development of musical polyphony), but have only relatively recently been put on an empirical footing in a classical series of studies by Van Noorden (1975, 1977). Although in this illustration two streams are formed when the pitch difference is relatively large, it is useful to keep in mind that streaming does not apply in an all-or-none fashion; there are intermediate conditions in which the stream will be more or less coherent.

Crucially for the arguments to be developed here, as the tones become separated into streams, the perception of order of the tones is very greatly changed. As illustrated schematically in Figure 3, as the difference in frequency between adjacent tone pairs increased modestly, linkages between events became stronger (see Figure 3A). With very great frequency differences between the tones, the linkages became realigned; this time, they bound those tones within a stream (see Figure 3B). Generally, it is the case that as the difference in pitch between alternating sounds is increased, they form two streams in which order information is preserved; however, judgments of order going across streams are severely impaired (Bregman & Campbell, 1971; see also Bregman, 1990). In the words of Bregman (1990), "the listener loses the finer details of the temporal relation between sounds that are in different streams and focuses instead on the temporal relations that exist among the elements of the same stream" (p. 143). We used this phenomenon in Experiment 3 to test the idea that the relation between distinctiveness and disruption is nonmonotonic and dependent on perceptual organization or streaming.

Experiment 3

In Experiment 3, we attempted to vary the degree of mismatch through a range of values that extended from the
relatively modest—at which level we may expect order to be preserved (and in conditions of deliberate listening a single, but varying, stream to be heard)—to the extreme, in which we expect the sound sequence to break up into two streams corresponding to conditions in which two unvarying sequences of sounds would be perceived. In this latter case, in terms of the changing-state hypothesis, little order was embodied. As we did this over a range of values, we expected disruption of serial recall first to rise, then to fall.

This prediction of nonlinearity was tested in Experiment 3 using the example just given of streaming by pitch difference. The use of change of pitch to explore nonmonotonicity is relatively straightforward technically; moreover, there is a substantial body of data relating to the phenomenal effects of pitch difference on streaming (see Bregman, 1990).

Our selection of pitch differences is informed by previous work in which the degree of streaming has been the main variable of interest. The critical value of pitch difference is the one at which typically one stream gives way to two; examination of data by Nootbooom, Brox, and de Rooij (1978) for pitch differences in sequences of speech, and data by Van Noorden (1975) for tones shows that this is somewhere on the order of 5 semitones (conveniently, data from both sets of studies are portrayed in Handel, 1989, p. 195). Using roughly comparable rates of presentation, in Experiment 3, we expected, therefore, that as we varied the pitch difference from 0 through 2, and 5 semitones, we would observe a monotonic deterioration in serial recall, but that, as the difference rose from 5 to 10 semitones, serial recall would improve significantly. We predicted, therefore, a quadratic function relating pitch difference and serial recall error.
In Experiment 3, we also compared these parametric effects in tones and pitch-shifted speech. We expected that the shape of the function would be the same in each case. The phenomenal effects of these pitch changes while in focal attention were already established in the work of Van Noorden (1975). Above 5 semitones it is expected that stream segregation will occur.

**Method**

**Participants.** Sixty-eight students volunteered for a small honorarium. All reported normal hearing and normal (or corrected-to-normal) vision. Participants were randomly assigned to one of two groups, one to a condition in which tones were to be used and the other to a condition in which speech stimuli were used.

**Materials.** Lists for serial recall comprised random orderings of the consonants f, k, l, m, n, r, s, t, and v; these were presented on the screen of a PowerMacintosh using SuperCard software (Version 2.5, Allegiant, San Diego, CA). The same items were used for each trial.

For the speech sequences, the vowel i was recorded in a male voice at a pitch of Ab2, and then digitally pitch shifted by 2 semitones (Bb2), 5 semitones (Db3), and 10 semitones (Gb3). All utterances were edited to 16-bit resolution at a sampling rate of 48 kHz, using Sound Designer II software (Degidesign, Menlo Park, CA). Each utterance lasted 100 ms and the interstimulus interval was 75 ms. Three sequences of alternating high and low pitch of the same vowel were constructed, separated by 2, 5, and 10 semitones, and one steady-state sequence (0 semitone separation). For the nonspeech sequences, a tone was constructed using PowerSynthesizer software (University of Sussex, Brighton, United Kingdom) of pitch E4, and then digitally pitch-shifted by 2 semitones (Gb4), 5 semitones (Ab4), and 10 semitones (Db5). All tokens were edited to 16-bit resolution at a sampling rate of 48 kHz, using Sound Designer II software. Just as for the speech, each stimulus lasted 100 ms and the interstimulus interval was 75 ms. Three sequences of alternating high and low pitch were constructed separated by 2, 5, or 10 semitones, and one steady-state sequence (0 semitone separation).

**Experimental design.** A mixed (within-between) design was used: Stimulus class was varied between subjects (two levels, speech and nonspeech); pitch separation was a within-subject variable. Four auditory conditions (four different levels of pitch separation) were randomly allocated to each trial with the restriction that all four conditions must have been presented before a condition could be repeated. In all, participants undertook 60 trials: 15 trials for each presentation condition. Before the experiment proper, participants were given two practice trials in silence.

**Procedure.** Sound was presented during both the period in which the to-be-remembered items were presented and the 10-s retention interval. Participants were tested individually in a single session lasting approximately 45 min. They were seated in a soundproof laboratory at about 0.5 m from the computer’s screen.

The general procedure was as before, with the exception that the sound playback was started slightly in advance of the presentation phase to permit time for any streaming effect to be perceptually established. During this 2-s pre-exposure period, a fixation point “+” appeared on the screen.

**Results and Discussion**

As before, responses were scored to a strict serial order criterion. The general form of the results is shown in Figure 4. The functions for speech and nonspeech stimuli took the same form; they were as predicted by the modified form of the changing-state hypothesis, namely, disruption increased with the degree of separation in pitch up to the 5-semitone level, but thereafter performance improved.

In a 2 (stimulus class: speech and nonspeech, a between-subjects variable) × 4 (frequency separation: 0, 2, 5, and 10 semitones, a within-subject variable) ANOVA, the effects of frequency separation were significant, $F(3, 198) = 12.98$, MSE = 0.13, $p < .001$, but the effect of stimulus class was not significant, $F(1, 66) = 1.33$, MSE = 22.84, $p > .05$, nor was the interaction of frequency separation and stimulus class significant, $F(3, 198) = 0.163$, MSE = 0.80, $p > .05$. As Figure 4 shows, there was a tendency for speech to be more disruptive generally than tones, and with sufficient power the difference might indeed be statistically significant.

Comparisons across different levels of frequency separation are of particular interest, the 5 and 10 semitone conditions being perhaps the most crucial. The increase in errors with frequency differences ranging from 0 to 2 semitones was significant, $F(1, 198) = 5.72$, $p < .05$, as was the increase in error as the frequency separation increased from 2 semitones to 5 semitones, $F(1, 198) = 14.13$, $p < .005$. Crucially, for the hypothesis under test, the decrease in errors between 5 and 10 semitones was significant, $F(1, 198) = 7.47$, $p < .01$. The nonmonotonic form of the function relating frequency difference and error was further substantiated by an analysis by orthogonal polynomials: The linear trend within the curve was significant, $F(1, 198) = 19.63$, $p < .0001$, and the quadratic term also was significant, $F(1, 198) = 12.14$, $p < .0005$. Notably, the level of disruption for speech at the 2-semitone level was the same as for nonspeech at the 5-semitone level (see Figure 4). Therefore, in

![Figure 4](image-url)

**Figure 4.** Experiment 3: Percentage of recall errors in relation to pitch separation shown for speech and nonspeech.
terms of numerical values of disruption, speech and tones were equipotential.

The results quite clearly demonstrate that the function relating pitch difference and degree of disruption was nonmonotonic; there was a point of inflection in the function above which the effect of a large pitch difference was beneficial. This is in line with the idea that two separate streams are formed when the differences in pitch are high, and that these streams contain less changing-state information because they are based on events whose pitches are unchanging. The magnitude of the improvement as the pitch difference went from 5 to 10 semitones, though statistically robust, did not produce performance at the level comparable to that found when a single steady-state sound was heard as irrelevant speech. This may be because the effect of having two steady-state streams is more disruptive than having one steady-state stream. More likely, as studies of the phenomenology of streaming suggest, the process of stream separation is relatively unstable, so that the two channels are not rendered consistently. In any case, whatever the precise level of performance in relation to the repeated-token conditions, analytically, the significant improvement in going from 5 to 10 semitones seems to suggest quite strongly that streaming processes were at work and that these modulated the disruptive effect of irrelevant sound.

One way of extending the results of Experiment 3 would be to observe the effect of introducing small changes in frequency within each stream. That is, in addition to the two streams formed from the alternation of high and low tones separated by 10 semitones, small changes (of less than 5 semitones), could be superimposed on the high and low streams in such a way as to reintroduce stimulus mismatch within stream, and hence once again increase disruption. Of course, Experiment 1 had a mixture of tones, but in that case the number of different tones was large and in random order, which meant that it was unlikely that different streams would form.

**General Discussion**

The results have contributed substantially to the development of the changing-state hypothesis. Hitherto, it had been assumed that the physical differences between successive stimuli would alone account for the degree of disruption. The results of the current series suggest a modification to this position in that the phenomena of streaming need also to be taken into account. One drawback of the changing-state hypothesis as it was previously conceived was that part of its reasoning was circular: Changing state was inferred from performance data in which an impairment had been produced and vice versa; there was no index of what specific conditions might produce the effect independent of the performance data. Within the current article, however, we have been able to bring to bear evidence that begins to overcome this difficulty.

In Experiment 3 we were able—on the basis of data from studies that pinpointed the level of pitch difference between alternating tones that produced fission (e.g., Van Noorden, 1975, 1977)—to predict the point of inflection in the function relating pitch difference to disruption of serial recall. This is buttressed by evidence that, when high and low tones are alternated, at the point where the pitch difference produces two streams, order relations between the high- and low-pitch streams become rather poor (Bregman & Campbell, 1971). There is corroborative evidence of a slightly different sort for Experiment 2, suggesting that highly discriminable acoustic sequences produce a rather poor representation of order (e.g., Warren & Obusek, 1972). Generally, the well-established literature on temporal order of sounds points to the fact that the order of events is well represented in sequences that show some change but that share some common characteristic, such as the voice in which the tokens are uttered (see Bregman, 1990, for an extensive discussion). That is, we now have some independent evidence that the conditions in which irrelevant sound is most disruptive are those in which, under different conditions of testing, the order of events is better preserved. This tends to verify a central tenet of the changing-state hypothesis, that the disruption by irrelevant sound is the result of the information about seriation that the irrelevant stream embodies. That is, the degree of disruption by an irrelevant sound sequence may be predicted from a separate estimate of memory for order of that sequence (specifically, when that sequence is learned deliberately).

The findings also have some precedent from the research in a paradigm in which the primary interest is the effect of interpolating different classes of sounds between two tones whose similarity has to be judged. A classical finding is that interpolated spoken digits (in which none are repeated) are less disruptive than interpolated tones (Deutsch, 1970). Jones, Macken, and Harries (1997) argued that this was because spoken digits formed a more coherent entity quite distinct from the tones whose similarity had to be judged. Indeed, by repeating a digit (rather than by having a random sequence) in the interpolated sequence, levels of recognition performance similar to those for tones were found. It is as if variation on a common carrier (the voice) is a particularly coherent entity, a suggestion that is reinforced by the finding that, by changing the fundamental of each digit in a sequence of varying digits, performance was again reduced to a level approaching that of interpolated tones. In sum, therefore, introducing large variation into a sequence threatens its integrity as a coherent entity, just as the results of Experiments 2 of the current series have shown.

The changing-state hypothesis is embodied within the object-oriented episodic record (O-OER), and it is the object feature of the model that proves useful in explicating the effects of distinctiveness in the irrelevant speech paradigm. That is, the formation of streams—in essence, temporally extended objects—is dependent on the physical mismatch between sounds in the sequence, and this in turn determines the degree of changing-state information in a sequence. In addition, the theory suggests, but does not specify fully, a mechanism that is common both to the organization of sound in auditory perception and to the representation of events in memory. The information relating to seriation that this process yields comes into conflict with that from the serial recall task.
The O-OER model embodies features that allow for the effects of streaming (see also Jones & Macken, 1995a, 1995b); the strength of the model, and in particular its embodiment in the changing-state hypothesis, is that it gives proper prominence to the effects of the integration of sound and its interplay with streaming. Moreover, the mechanism of interference is also specified, because the effects of irrelevant sound seem to be most marked in memory tasks that involve the preservation of serial order. Each of these elements reinforces the generalization that interference by irrelevant sound is a result of a conflict of common processes, namely ones of seriation (see Jones et al., 1996). A single explanatory concept accounts for a range of effects: the complex pattern of interference with different types of sound and the pattern of sensitivity of different tasks (see Beaman & Jones, 1997, 1998; Henson, Burgess, & Hitch, 1997; but see LeCompte, 1996; LeCompte et al., 1997).

The current series has extended the empirical base of the changing-state hypothesis and, at the same time, concerns about elements of circularity in the concept of changing state and its relation to seriation have been addressed.

References


Received November 3, 1997
Revision received May 11, 1998 -
Accepted May 22, 1998