Does Auditory Streaming Require Attention? Evidence From Attentional Selectivity in Short-Term Memory

William J. Macken, Sébastien Tremblay, Robert J. Houghton, Alastair P. Nicholls, and Dylan M. Jones
Cardiff University

R. P. Carlyon, R. Cusack, J. M. Foxton, and I. H. Robertson (2001) have argued that attention is crucial for auditory streaming. The authors review R. P. Carlyon et al.’s (2001) arguments and suggest that a pertinent literature, the irrelevant sound paradigm—demonstrating preattentive auditory streaming—has been overlooked. In illustration of this alternative approach, the authors include a novel single experiment demonstrating the impact of preattentive auditory streaming on short-term serial memory. It is concluded that R. P. Carlyon et al.’s (2001) results do not definitively demonstrate that auditory streaming processes are dependent on attention; indeed, they are compatible with alternative accounts of the relationship between perceptual organization and attention.

Some 4 decades of experimentation have served to delineate extensively the functional properties of what has become known as auditory scene analysis or auditory streaming (e.g., Bregman, 1990). This function of auditory perception refers to the decomposition of the specific pattern of acoustic stimulation arriving at the two ears into its constituent elements in terms of their environmental origin. Thus, a representation of the organism’s environment is derived from acoustic information in a manner that is well characterized by a process embodying Gestalt principles applied to aspects of a given stimulus, such as frequency, location, and timbre. This Gestalt characterization of auditory scene analysis is not generally controversial (e.g., Handel, 1989). However, whether or not such auditory scene analysis may take place outside the focus of attention remains an issue of debate. The functional pattern of auditory scene analysis has been mapped almost exclusively on the basis of methodologies that require focal attention to, and subjective judgment of, critical stimuli. As such, the corpus of data on auditory scene analysis do not necessarily speak to the role of attention in auditory streaming because almost all examples of such streaming phenomena derive from tasks that involve focal attentional processing.

Although there have been indirect attempts to address this issue in the past (e.g., Bregman & Rudnicky, 1975), little direct testing of the question has taken place; a shortfall that the studies of Carlyon, Cusack, Foxton, and Robertson (2001) set out to redress. As such, the Carlyon et al. article marks an important development in the study of the role of attention in auditory stream formation. In the current article, we summarize the findings of Carlyon et al., before reviewing findings from a paradigm not considered by Carlyon et al., but which nonetheless is highly pertinent to the assessment of the role of attention in auditory streaming. Unlike the conclusion of Carlyon et al., the review of these studies raises the possibility that auditory scene analysis may occur preattentively. Critically, the dependent measure of streaming in these studies does not rely on any attention to, or judgment of, the auditory stimulus. Further, a novel empirical demonstration of such preattentive auditory streaming is reported, again avoiding any contaminating effects of deliberate focal attention or decision processes. Finally, in this article, we consider the aspects of methodology and conceptualization of the issue that may have led to these disparate conclusions.

Studies of Carlyon et al. (2001)

Carlyon et al. (2001) presented two pairs of experiments, the first of which established preliminary bases for arguing that stream segregation is an attention-demanding process. The second pair of experiments applied to the study of hemispheric neglect with participants drawn from a clinical population. In this article, we restrict our comments and reply to the first pair of experiments, rather than dealing with neuropsychological aspects of streaming.

To study the role of attention in streaming, Carlyon et al. (2001) focused on a particular temporal characteristic of stream segregation: namely, that under certain conditions, sequential segregation may take some seconds to “build up.” In other words, a sequence that is initially perceived as a single stream may, over time, become perceived as two streams (see, e.g., Bregman, 1990). In both the experiments under consideration here, Carlyon et al. investigated the effect of attentional manipulations on the time course of “number-of-stream” judgments on tone sequences of the
type ABA–ABA–ABA, in which A and B are tones of different frequencies. At low levels of frequency separation and/or slow rates of presentation, a single stream with a “galloping” rhythm is heard, whereas with large frequency separations and/or fast rates of presentation, two streams are perceived. Noteworthy here is that the paradigm adopted for investigating the role of attention in streaming involves (as has the research examining other functional aspects of streaming) subjective judgments of stimuli that are within the focus of attention.

In Experiment 1 of Carlyon et al. (2001) tone sequences were presented to one ear for 21 s, and participants were required to make number-of-stream judgments for sequences of different A–B frequency separations. In the divided attention condition, participants were also required to respond, for the first 10 s of the trial, to a sequence of noise bursts presented to the other ear. Each noise burst either linearly increased in amplitude (approaching) or decreased in amplitude (departing), and the participant was required to discriminate each burst in the sequence accordingly. After 10 s of this task, participants were then required to make number-of-stream judgments on the tone sequence that had also been running simultaneously with the noise sequence. This divided attention condition was contrasted with two control conditions. In one, the tone and noise sequences were presented together, but no response was required to the noise burst sequence, whereas in the second, only the tone sequence was presented and the participant began making number-of-stream judgments from the beginning of the tone sequence. After 10 s, stream segregation had built up to a lesser degree in the divided attention condition than in either of the control conditions, and the time course of stream buildup over this last portion of the tone sequence was akin to that found at the beginning of the tone sequence in those conditions in which attention was focused on that sequence from the beginning.

In Experiment 2, the secondary attentional task was carried out on the tone sequence itself, rather than on a separate sequence. In this case, the tones in the sequence were amplitude modulated at either a fast rate (16 Hz) or a slow rate (8 Hz). Within the sequence, AM rate changed approximately every four tone bursts. In one condition, participants were required, for the first 10 s, to detect changes in the AM rate, after which they began their number-of-stream judgments. In the other condition, number-of-stream judgments were required from the beginning of the sequence. In this experiment, no difference in stream buildup after 10 s was found between the conditions, attributable, according to Carlyon et al. (2001), to the fact that in both conditions, participants were attending to the tone sequence throughout its duration.

As noted above, this conclusion is based on results from a paradigm involving focal attention to the critical stimulus. Therefore, it is worth being cautious about the conclusions of Carlyon at el. (2001) at this point, given that attentional set may itself influence whether or not stream segregation is perceived (e.g., van Noorden, 1975). Further, tasks that involve subjective judgments of critical stimuli will necessarily also involve decision processes beyond those processes that may be involved in stream segregation per se. Below, we review evidence on the role of attention in streaming from a paradigm, the irrelevant sound effect (ISE), which does not involve focal attention to the auditory stimulus but which nonetheless may inform the debate as to whether or not streaming requires attention.

By its mere presence, task-irrelevant background sound disrupts serial recall for visually presented items (e.g., Colle & Welsh, 1976; Salamé & Baddeley, 1982). Typically in this paradigm, participants are presented sequentially with a series of items (such as consonants or digits) on the screen of a computer. Normally, from seven to nine items are presented, typically at a rate of around 1 item per second. After presentation of the lists, there is normally a retention period of up to 10 s, followed by instructions to the participant to recall the items in their initial presentation order. Some such trials are accompanied by irrelevant sound; others take place in quiet. Despite explicit instructions to ignore the sound, coupled with the assurance that there will be no subsequent test of its contents, recall is markedly impaired by the irrelevant sound.

Critical aspects of this methodology suggest strongly that the background sound genuinely remains unattended. Generally, three main types of safeguard need to be embodied in any selective attention experiment using unattended stimuli (ten Hoopen, 1996). First, the extent to which a primary task is demanding in terms of processing resources must not be so low as to allow the risk of participants’ voluntary switch of attention toward the irrelevant source. Second, the discrimination between relevant and irrelevant material should be easy. Third, shifts of attention are more likely in settings in which participants are asked to report changes in the irrelevant material. In the typical irrelevant sound paradigm, all three criteria are met. Serial recall of supraspan lists is a demanding primary task, the to-be-remembered items are presented visually, and the to-be-ignored items are presented in the auditory modality, respectively, and participants are assured they will never be tested on the contents of the irrelevant material nor is any such test ever administered. Taken together, the scope and extent of these safeguards suggest that the disruptive effect of irrelevant sound is not the result of deliberate, focal processing.

As well as these methodological safeguards, there are aspects of ISE that further suggest that deliberate, attentional processing does not occur. For example, it might be reasonable to expect that making the sound more interesting would increase the risk of attention being deliberately diverted to it, but there is no evidence that this is the case. First, increasing the level of meaning in the irrelevant sound fails to increase interference (e.g., contrasting reversed with forward speech; Jones, Miles, & Page, 1990). Second, making the sequence of tokens composing the irrelevant sequence more predictable also does not reduce its impact: Thus, repeating the sequence ABCDEF is no more disruptive than if the tokens are presented in a random order (Jones, Madden, & Miles, 1992; Tremblay & Jones, 1998). Third, greater variability in terms of the number of different events in a sequence does not promote greater disruption. When the effect of the number of different tokens in the irrelevant sequence is manipulated systematically, disruption increases markedly as the number of tokens increases from one to two but increases very little as the set size increases above that (Tremblay & Jones, 1998). If the disruptive effect of irrelevant sound were due, at least in part, to its ability to divert attention from the focal task, then we would expect such manipulations to affect the level of disruption, but it appears that they do not.

Further, if there was a tendency for the participant to listen to the irrelevant sound it might be expected that this would diminish as
the participants became more familiar with the material. However, this does not appear to happen; the magnitude of ISE does not diminish over trials (e.g., Hellbrück, Kuwano, & Namba, 1996; Jones, Macken, & Mosdell, 1997), even when they are blocked by condition or when the sessions are days apart (Hellbrück et al., 1996). Although it may be impossible to fully rule out the chance that participants pay some momentary attention to the irrelevant sound, the aspects of methodology and evidence described above at least provide a case for using the irrelevant sound paradigm in the study of the role of attention in auditory scene analysis.

The last decade has seen a proliferation of research on this paradigm (see Jones, 1999, for a review) because it may illuminate the nature of obligatory processing of auditory stimuli and the way in which this process interferes with short-term memory. The weight of evidence (see, e.g., Macken, Tremblay, Alford, & Jones, 1999) indicates that the specific identity of the sound is not important to its disruptive capacity (e.g., Jones & Macken, 1993) nor does the degree of similarity of content between the background sound and the material in the focal task play a role (Macken & Jones, 1995). So, the ISE is found with stimuli as diverse as narrative speech and sequences of tone bursts. What is crucial to the disruptive effect is that the sound must be segmentable into discrete, sequential entities as in the case of tones, syllables, and words (Jones, Macken, & Murray, 1993). Each segmented entity must be different from that which preceded it: Sounds that are either perceptually continuous or involve repetition of a single token produce little or no disruption of recall (e.g., Hellbrück et al., 1996). Although it may be impossible to fully rule out the chance that participants pay some momentary attention to the irrelevant sound, this does not appear to happen; the magnitude of ISE does not diminish over trials (e.g., Hellbrück et al., 1996; Kilcher & Hellbrück, 1996; Kilcher et al., 1997), thus diminishing the impact on order memory in the focal task. A further illustration of the modulating role of streaming by location in the ISE paradigm is provided by studies of the effect of babble speech: an acoustic signal comprising many different voices (e.g., Kilcher & Hellbrück, 1996; Jones & Macken, 1995a).

As more voices are added to a stimulus, the disruptive capacity of that sound is reduced. This effect can be accounted for in terms of a cumulative masking of cues to segmentation (such as the abrupt frequency and/or amplitude changes at boundaries of syllables or phonemes) by acoustic components of the other voices, thus diminishing the degree of changing state within the stimulus. A significant reduction in disruption can be demonstrated with babble comprising as few as six voices (Jones & Macken, 1995a). It is important to note that, this diminution in disruption only occurs if the compound signal emanates from a single spatial location. However, if these six voices are presented each from a separate loudspeaker arranged around the participant’s head, then the level of disruption is increased relative to conditions where the six voices are presented from a single spatial location (Jones & Macken, 1995a, Experiment 4). This strongly suggests that, despite being unattended, the auditory stimulation arriving at the two ears is being decomposed on the basis of critical acoustic attributes into its constituent voices, and each voice then constitutes a discrete perceptual entity that can be processed as if it were presented on its own. In other words, having streamed out the separate voices from the compound stimulus, each voice can then be segmented and therefore register changing state. Consequently, its disruptive potency is restored. Again, this result indicates that the process of stream segregation is being carried out in the absence of focal attention.
Example 2: Auditory Restoration

If a segment of an auditory stimulus is masked by some other auditory event, under certain conditions, the former auditory stimulus is perceived as continuing “behind” the masking stimulus. Further, if a segment is actually excised from an ongoing stimulus, such as an FM glide, and is replaced with another stimulus that would have the capacity to mask the glide, then the auditory system perceptually restores the missing segment, giving rise to the illusory perception of continuity. Such a phenomenon has been demonstrated using both speech and nonspeech sounds (see, e.g., Warren, 1970; Dannenbring, 1976). Given that the methodology typically used here involves focal attention to, and subjective judgment of, such stimuli, clearly here too the precise role of attention may be questioned. Again, the ISE paradigm provides evidence that focal attention is not required for this aspect of auditory scene analysis to be accomplished.

As noted above, perceptually continuous auditory stimuli do not disrupt serial recall appreciably (Jones et al., 1993). So, for example, a continuous, randomly varying FM glide does not interfere with order memory. However, if such a glide is regularly interrupted by short periods of silence, thus creating a segmentable and changing sequence, disruption does occur. The critical comparison from Jones et al. (1993) in the present context is between two interrupted-glide conditions. In one, the silent interruptions were filled with a noise burst conforming to the amplitude and frequency characteristics that would be required to mask the stimulus and, therefore, that would induce auditory restoration. In the other condition, the acoustic properties of the noise burst fell outside this range. When these two stimuli were presented as irrelevant sound, the former behaved like acoustically continuous glides in that they did not interfere with recall, whereas the latter behaved like interrupted glides in that they interfered with recall (Jones et al., 1993; Experiment 3), further implicating the preattentive analysis of the auditory scene.

Example 3: Streaming by Pitch

The disruptive effects of irrelevant sound have also been shown to be modulated by perceptual organization on the basis of pitch. For a given rate of presentation, the phenomenal percept of a sequence of tones alternating in frequency will change from one of a single stream alternating in pitch to one of two streams, each of repeated pitch, once the frequency separation between the two tones reaches a critical level (the actual change in frequency required to produce this fission depends on the rate of presentation of the tones; van Noorden, 1975). In the context of the ISE paradigm, this process constitutes a transformation from one changing-state stream (a disruptive one) to two steady-state (non-disruptive) streams (see Figure 2). Jones et al. (1999) examined the effects on recall of sequences of tones as the change in frequency from tone to tone increased. Tones were presented at a rate of nearly six items per second (100 ms on, 75 ms off), while the participants were engaged in the typical irrelevant sound procedure. The effect on disruption of increasing the frequency separation in alternating tones from 0 (Ab2) to 2 (Ab2–Bb2) to 5 (Ab2–Db3) to 10 (Ab2–Gb3) semitones was assessed. For the particular rate of presentation used, the fission threshold would be expected to occur at around 5-semitone separations. For changes in frequency up to this level, disruption of recall increased as the level of changing-state increased (see also Jones et al., 2000). However, recall performance improved between the 5- and 10-semitone separation conditions, indicating again that despite being unattended, the auditory sequence was processed either as a single, alternating entity or as two unchanging entities, depending on frequency separation, rather than as a single, fused percept. If the latter had been the case, there should not have been a drop in interference in the 10-semitone condition because it also would constitute a single, changing-state stimulus.

In this section, we have reviewed the behavior in the ISE paradigm of three aspects of auditory scene analysis: streaming by location, auditory restoration, and streaming by pitch. In each case, the results indicate that such perceptual organization is taking place outside the focus of attention of the participant. Below, we further bolster this argument by presenting a novel empirical test, this time using another cue to auditory stream segregation—that of presentation rate.

Present Experiment: Preattentive Streaming by Rate

At a functional level, auditory scene analysis capitalizes on critical acoustic properties of an auditory input in order to organize that input into a representation of the various events present in the
auditory environment. In the previous section, we have described how the preattentive processing of a number of such properties is evidenced in terms of their modulating role on disruption of recall by background sound. This pattern is a general one, in that the same conclusions about such preattentive processing can be sustained across a range of stimuli and a range of cues to streaming. In this section, we set out to provide yet further evidence for the preattentive processing of the auditory scene by testing the impact on order memory of a previously unexamined cue, namely, rate.

The reciprocal relationship between frequency and timing as cues to stream segregation has been extensively analyzed (e.g., Bregman, 1990; van Noorden, 1975). What such studies show is that the fission threshold (in which a single-stream percept changes to a dual-stream percept) is a function of both the frequency separation of alternate tone bursts and the rate of presentation of the tone bursts. So, a tone sequence within which a certain frequency change occurs may give rise to a single- or a dual-stream percept, depending on the rate at which the tone bursts occur. Generally, the greater the frequency separation between alternate tones, the greater the rate of presentation needs to be in order to induce fission. Similarly, relatively small frequency changes may still give rise to stream segregation if the rate is sufficiently high.

In this experiment, frequency separation is held constant while the rate of presentation of tone bursts is modulated. According to the account of disruption in serial recall described above, an increase in changing state leads to an increase in disruption but only to the extent that such changing state occurs within a particular perceptual stream. For example, the results of Jones et al. (1999) indicated that disruption of recall increases with frequency separation of alternate tones, up to the point at which frequency separation subsumes stream segregation. Bridges and Jones (1996) have also shown that an increasing rate of change leads to increasing disruption of recall. Again, such a phenomenon accords with the idea that the degree of changing state within a background sound positively determines the degree of disruption of serial recall. Accordingly, we would predict an increase in disruption of recall with an increased rate of presentation of tones but only up to the point at which the frequency separation-by-rate relationship tends to subsume a fissile percept. At this rate, as in the streaming by pitch example (Jones et al., 1999) described above, the changing state sequence becomes two steady-state sequences, and thereby, its disruptive potency is reduced. Preattentive streaming by rate is a necessary condition for this pattern of interference to occur; a single, fused representation of the auditory stimulus would not produce such a nonmonotonic relationship between rate and disruption.

Method

Participants

Forty volunteers, each of whom reported normal hearing, were recruited among students at Cardiff University, Cardiff, United Kingdom. They received a small honorarium for their participation.

Materials

Two pure tones were generated, one lower by seven semitones in pitch than the other (349 Hz). Each tone was 100 ms long (with a rise time of 15 ms). The tone sequences followed the “galloping triplets” form of van Noorden (1975), being arranged in triplets of low–high–low tones. From these stimuli, three sets of irrelevant sound, differing in their rate of presentation, were created by varying the duration of interval between offset and onset of the tone bursts: (a) 10 ms, (b) 100 ms, and (c) 230 ms. An additional sequence, consisting of a single tone played repeatedly at a regular rate (100 ms on, 100 ms off) was also used, which acted as a control condition. In half of these trials, the repeated tone was at the low frequency; in the other half, the repeated tone was at the high frequency. Lists of items to be recalled were presented visually on the screen of a Power Macintosh (Model 7200/90) computer. The sequences were constructed from random orderings of the consonants F, K, L, M, Q, R, Y, S, and T, with the constraint that a consonant could not be presented in the same serial position in two consecutive lists. The consonants were displayed individually at a rate of one per second.

Design

Five auditory conditions (low rate, medium rate, high rate, repeated tone control, and quiet control) were randomly allocated to each trial with the
constraint that all five conditions must have been presented before a condition could be repeated. There were 15 trials per condition and 75 trials in all.

Procedure

Participants were tested individually, seated in a soundproof laboratory, at a distance of approximately 50 cm from a computer screen. At the outset, participants were given standard instructions on the screen, informing them of the requirements for serial recall and instructing them to ignore any sounds they might hear. Participants were also informed that the sounds would not contain messages and that they would not be tested on their contents. The experimental trials were preceded by a short practice session of three trials in quiet. The participant initiated each trial by using a mouse button to click on a SuperCard virtual button on the screen. This started the presentation of the stimulus sequence in which the consonants were displayed individually at a rate of one per second (on for 0.8 s, off for 2 s). When nine consonants had been presented, the word wait flashed on the screen for 10 s, during which participants were expected to rehearse covertly. The word recall was then displayed to prompt participants to write down their response, in the order in which they were presented, on a response blank. The irrelevant sound was presented through headphones throughout presentation and rehearsal phases and was switched off automatically during recall. In all, the experiment took approximately 50 min to complete.

Results and Discussion

Figure 3 illustrates the percentage serial recall error pooled over serial positions for the five auditory conditions. Visual inspection suggests that the relation between the token-dose (interval duration) and degree of disruption is not monotonic. A repeated-measure analysis of variance (ANOVA) was carried out on the error data with serial position (nine levels) and conditions (five levels). Both main effects of auditory condition, $F(4, 156) = 6.25, MSE = 7.21, p < .001$, and of serial position, $F(8, 312) = 39.95, MSE = 13.74, p < .001$, were significant. The interaction between auditory condition and serial position was not significant, $F(32, 1248) = 0.85, MSE = 2.12, p = .62$. A set of pairwise planned comparisons was performed on the main effect of auditory condition. The alpha significance level was set at .0125 as a Bonferroni correction for making four comparisons. The repeated token sequence did not produce significantly more disruption than quiet, $F(1, 156) = 0.04, p = .910$. The changing sequence played at a low rate was more disruptive than the repeated token sequence but did not reach the level of significance that was set, $F(1, 156) = 4.28, p = .040$. Comparisons across the three different levels of token-dose are of particular interest. The increase in error from low to medium dose approached significance, $F(1, 156) = 4.00, p = .047$, but most important, there was a marked and significant decrease in errors between medium and high dose, $F(1, 156) = 12.93, p < .001$. Indeed, the level of error in the high-rate condition approximated that of the repeated-token condition. A trend analysis revealed a significant quadratic function relating the rate of presentation (low–medium–high) and the error in serial recall, $F(1, 156) = 7.82, p < .001$.

Clearly, the form of the function relating presentation rate and disruption corresponds to that predicted if it were the case that preattentive auditory streaming had taken place. Rate of presentation of tones increased the degree of disruption up to a point at which the changing single stream was split into two unchanging streams, thereby reducing disruption. This pattern of results converges with that of Jones et al. (1999) in which pitch separation was varied for constant rate of presentation of tone bursts in showing that the relationship between change and disruption is critically modulated by the perceptual organization of the unattended auditory environment. In all respects other than rate of presentation, the auditory stimuli used in the present experiment were identical. However, the nonmonotonic relationship between their disruptive capacity and their rate makes sense only if we allow for the segregation of the auditory scene to take place outside the focus of attention, as it has been demonstrated within the focus of attention.

Conclusion

In this article, we have set out to address the critical theoretical question: Does auditory streaming require attention? Until recently, the streaming literature has not provided a resolution of this issue. Carlyon et al. (2001) have provided a timely attempt to directly test the role of attention in stream segregation and conclude that attention is crucial for the buildup of auditory streaming. However, the method that Carlyon et al. used was one that involved focal attention to a critical stimulus and subjective judgment of whether or not that stimulus was streamed or unstreamed. The methodology involved in studying ISE does not suffer from such problems because participants are neither required to attend to the sound nor to make any judgment about it. Rather, the index of how the unattended sound is processed is given by its disruptive impact on a separate, focal task. Further, the methodology conforms to criteria that are necessary to provide face validity to the claim that a given stimulus is genuinely unattended.

There are various ways in which the current results could be reconciled with those of Carlyon et al. (2001). One possibility is that participants do to some extent attend to the irrelevant sound, although throughout this review we have stressed a number of
good reasons why we do not think this is likely to be the case (see also ten Hoopen, 1996). Another possibility is that the short-term memory task used in the irrelevant speech paradigm is less attentionally demanding and hence less disruptive to the perceptual organization of auditory stimuli than the auditory discrimination tasks used by Carlyon et al. (2001). It is true that the attentional demands in a serial order short-term memory experiment are not constant over time; in the early stages of presentation and encoding, the participant only needs to retain as many items as have at that point been presented. If the effects of irrelevant sound can only been found during this phase of the experiment, one may suspect that at this point participants may have had spare attentional capacity to devote to listening (despite instructions to the contrary) to the irrelevant auditory material (and thus allow the onset of the attentional buildup of streaming). However, as well as being disruptive when occurring simultaneously with presentation of the memory items, irrelevant speech has also been shown to be disruptive when its presentation is restricted to the retention phase of the experiment, after the complete list has been presented, and as such, where the attentional requirement to perform the primary task is uniformly high across the interval. This indicates that the irrelevant sound has its effect at a central or memorial level rather than at encoding; moreover, the effect is most evident at the point when the participant is experiencing the greatest burden of processing. For example, when sound is presented only to coincide with presentation of the initial few memory items, it does not actually interfere with memory for those items. However, when sound presentation is restricted to that period of time immediately after the complete memory list has been presented, significant interference with those early list items does occur (Macken, Modell, & Jones, 1999). If it were the case that participants were dividing their attention during the early stage of list presentation, then one might expect to see that divided attention manifested in some degree of performance decrement. However, the same level of recall performance is observed when sound presentation is restricted to this early stage as is found under conditions in which no sound at all is presented. In other words, at that point in time when the primary task might be of sufficiently low load to allow the participant to attend to the sound, the presence of the sound has no impact whatsoever on performance on that primary task. We would argue that this strongly suggests that participants do not choose to attend to the irrelevant sound, even early on in the task, and that, therefore, the streaming does occur outside the focus of attention.

One final argument in this vein is that, precisely because the primary task load is so low during the early stage of list presentation, deliberately listening to the irrelevant sound would incur no cost on primary task performance, and so participants may well listen to the sound at an early point in time, but this divided attention situation has no behavioral consequences. From this point of view, it may be that this low-task-load situation allows participants to build up a streamed percept of the irrelevant sound in the examples described above by attending to the sound at the beginning of a trial. When the task load becomes sufficiently high, participants then ignore the sound, but the necessary streaming has already taken place, not necessarily within the focus of attention, but nonetheless in a situation in which some attentional resources are being devoted to processing the irrelevant sound. That this is also not the case is suggested by data from Jones and Macken (1995b) in which streaming by location (as described above) was examined. Figure 4 presents results from two such experiments (Jones & Macken, 1995b, Experiments 1a and 1b). In one instance, irrelevant sound was presented during both the presentation and retention phases of the trial. As can be seen (Figure 4, left panel), streamed, stereophonic presentation of three syllable sequences produces less disruption than monophonic presentation. If participants do listen to the sound when the primary task load is low, then the fact that streaming occurs here is not evidence that it is taking place without attention. However, this pattern of results also obtains when the irrelevant sound only occurs during the retention phase of the trial, after the complete list has been presented, and as such, where the attentional requirement to perform the primary task is uniformly high across the interval. This indicates that the irrelevant sound has its effect at a central or memorial level rather than at encoding; moreover, the effect is most evident at the point when the participant is experiencing the greatest burden of processing. For example, when sound is presented only to coincide with presentation of the initial few memory items, it does not actually interfere with memory for those items. However, when sound presentation is restricted to that period of time immediately after the complete memory list has been presented, significant interference with those early list items does occur (Macken, Modell, & Jones, 1999). If it were the case that participants were dividing their attention during the early stage of list presentation, then one might expect to see that divided attention manifested in some degree of performance decrement. However, the same level of recall performance is observed when sound presentation is restricted to this early stage as is found under conditions in which no sound at all is presented. In other words, at that point in time when the primary task might be of sufficiently low load to allow the participant to attend to the sound, the presence of the sound has no impact whatsoever on performance on that primary task. We would argue that this strongly suggests that participants do not choose to attend to the irrelevant sound, even early on in the task, and that, therefore, the streaming does occur outside the focus of attention.

One final argument in this vein is that, precisely because the primary task load is so low during the early stage of list presentation, deliberately listening to the irrelevant sound would incur no cost on primary task performance, and so participants may well listen to the sound at an early point in time, but this divided attention situation has no behavioral consequences. From this point of view, it may be that this low-task-load situation allows participants to build up a streamed percept of the irrelevant sound in the examples described above by attending to the sound at the beginning of a trial. When the task load becomes sufficiently high, participants then ignore the sound, but the necessary streaming has already taken place, not necessarily within the focus of attention, but nonetheless in a situation in which some attentional resources are being devoted to processing the irrelevant sound. That this is also not the case is suggested by data from Jones and Macken (1995b) in which streaming by location (as described above) was examined. Figure 4 presents results from two such experiments (Jones & Macken, 1995b, Experiments 1a and 1b). In one instance, irrelevant sound was presented during both the presentation and retention phases of the trial. As can be seen (Figure 4, left panel), streamed, stereophonic presentation of three syllable sequences produces less disruption than monophonic presentation. If participants do listen to the sound when the primary task load is low, then the fact that streaming occurs here is not evidence that it is taking place without attention. However, this pattern of results also obtains when the irrelevant sound only occurs during the retention phase of the trial, after the complete memory list has been presented, and as such, where the attentional requirement to perform the primary task is uniformly high across the interval. This indicates that the irrelevant sound has its effect at a central or memorial level rather than at encoding; moreover, the effect is most evident at the point when the participant is experiencing the greatest burden of processing. For example, when sound is presented only to coincide with presentation of the initial few memory items, it does not actually interfere with memory for those items. However, when sound presentation is restricted to that period of time immediately after the complete memory list has been presented, significant interference with those early list items does occur (Macken, Modell, & Jones, 1999). If it were the case that participants were dividing their attention during the early stage of list presentation, then one might expect to see that divided attention manifested in some degree of performance decrement. However, the same level of recall performance is observed when sound presentation is restricted to this early stage as is found under conditions in which no sound at all is presented. In other words, at that point in time when the primary task might be of sufficiently low load to allow the participant to attend to the sound, the presence of the sound has no impact whatsoever on performance on that primary task. We would argue that this strongly suggests that participants do not choose to attend to the irrelevant sound, even early on in the task, and that, therefore, the streaming does occur outside the focus of attention.
interval (Figure 4, right panel). At this stage, the memory load in the primary task is at a maximum, but the pattern of disruption is essentially identical to that in Figure 4, left panel. When the acoustic conditions exist for a changing-state sequence to be streamed to create a number of steady-state percepts, disruption decreases, whether or not opportunities for deliberate listening exist. Although this evidence relates to streaming by location while the Carlyon et al. (2001) results relate to streaming by frequency, this evidence nonetheless suggests that the streaming effects observed within the irrelevant sound paradigm are not due to deliberate processing of the irrelevant sound.

Of course, this does not necessarily mean that the stream segregation that takes place without the focus of attention is identical in all its functional aspects to that which takes place within the focus of attention. Nor does it mean that different “amounts” of attention may be required to achieve different aspects of stream segregation. From this perspective, the evidence from the irrelevant sound paradigm does not necessarily contradict the conclusions of Carlyon et al. (2001), but it does qualify those conclusions by indicating that at least some aspects of auditory stream segregation do not require focused attention.

Alternatively, a fundamental methodological difference between the results from the ISE paradigm and those of Carlyon et al. (2001) may account for the discrepant conclusions. Specifically, given that the dependent variable in the Carlyon et al. experiments is subjective in nature, their attentional manipulation may have affected participants’ behavior at a level other than that at which the underlying streaming processes occur. For example, van Noorden (1975) showed that, within constraints, participants could select at will whether they perceived the type of ABA sequence used by Carlyon et al. as one or two streams. It is not inconceivable that the attentional tasks in Carlyon et al. had an influence on such a selection process. Alternatively, the attentional tasks may have affected the transfer of information from preattentive sensory representations through to consciousness, possibly by impairing their encoding in memory (see Moore & Egeth, 1997, for a similar account in visual perceptual organization). On the other hand, the ISE paradigm, in not requiring an attention to or judgment of the auditory stimulus, is immune from such concerns.

A full explanation of the difference in outcome between our results and those of Carlyon et al. (2001) will hopefully emerge from continued research into this important and now controversial research issue. What does seem clear, from both the review of the role of streaming processes in the modulation of disruption of recall by background sound, and from the new results presented here, is that, when critical methodological issues are addressed, auditory streaming can occur in the absence of focal attention.

References


Received February 1, 2001
Revision received April 9, 2002
Accepted May 16, 2002