

# Indispensable Benefits and Unavoidable Costs of Unattended Sound for Cognitive Functioning

R.W. Hughes and D.M. Jones

*School of Psychology, Cardiff University*

**Critical to survival, and also to the organism's efficient management of the flow of information in the brain, is attentional selectivity; the ability to select one source of information to guide action whilst ignoring others that are irrelevant to the current behavioural goal. But such selectivity is not merely the inclusion of the relevant information and the complete neglect of irrelevant information. We discuss in this paper the way that all sound is processed in an obligatory fashion - whether relevant or irrelevant - and discuss the fate of sound in the case when it is irrelevant to the immediate mental task. Using the so-called irrelevant sound paradigm we show that unattended information is both registered and organised. This obligatory process of organisation compromises the efficiency of particular types of mental activity. We discuss how such interference comes about but the key emphasis is upon the possible beneficial effects of such processing-of-the-irrelevant, in allowing the switching of attention to be more facile and intelligent and in allowing the accumulation of evidence about statistical regularities in the auditory world (such as those helpful to the efficient perception, acquisition and use of language). In sum, we describe how purposeful processing based on directed attention is in a state of tension with the obligatory, automatic processing of the unattended. One of the consequences of this tension is typically manifested in auditory distraction, but the benefits of processing of the attended may considerably outweigh this disadvantage.**

*Keywords: unattended sound, irrelevant sound effect, auditory distraction, auditory attention, cognitive functioning*

## **Introduction**

The nervous system is constantly subject to a barrage of information entering via the various senses. Selectivity, that capacity to pick out from this melange, the subset of events that are most relevant for current behavioural goals, is essential to successful and purposeful prosecution of those goals (the phenomenon of *selective attention*; see e.g., Allport, 1989, 1993; Pashler, 1998; Styles, 1997). That is, as organisms, we try to attend to the relevant source of information (e.g., the words on this page) and ignore irrelevant sources (e.g., the sounds emanating from traffic outside perhaps) so that our behaviour (e.g., reading) can be exclusively controlled by the task-relevant information.

A prerequisite for being able to pick out the relevant subset of information is that the

incoming mixture of sensory information must be partitioned according to the different objects in the environment. The process of partitioning not only generates a set of candidate objects from which the relevant object can be picked out, but also, given that what is relevant is of course a dynamically changing property of the organism's interaction with the world, it allows attention to be shifted adroitly to other objects. This means that the process of selection not only involves picking out the relevant object but also involves minimising the potential disturbance or 'crosstalk' deriving from the other candidate objects.

Very early in the history of our understanding of human cognition it became clear that events that were not attended were to some degree

processed. Indeed, it is now clear that information that is being ignored (i.e., not being attended) is being processed to the extent that it is being actively organised by the perceptual system. It is usual to make a distinction therefore between processing that is attentive, that is, part of the deliberate, purposeful and goal-directed action of the individual, of which that individual can be said to be aware, and that which is preattentive, subject to rather less elaborate processing, possibly only to the level of physical organisation, of which the individual is not necessarily aware. This preattentive analysis leaves the organism in a state of perpetual tension because the integrity of attentional engagement on the task-relevant information is now compromised to some degree by the fact that the preattentively formed objects will tend to compete for the control of action.

The key theme of the current paper is that this processing-of-the-unattended has a functional purpose; it is not merely the side-effect of a faulty selectivity, or some imperfection in the grand design of the nervous system. For the purpose of survival and adaptation, the tension produced by the need to engage purposefully on a subset of information whilst simultaneously organising goal-irrelevant (and therefore unattended) information is a healthy and indispensable one. Representing unattended information to the level of objects which then compete for the control of action serves a number of key functions. Because the organism already knows (in a relatively crude sense at least) 'what is out there', when new or other events become more relevant it allows for the flexible and swift re-allocation of the focus of selective attention onto previously unattended sources of information (e.g., Houghton and Tipper, 1994). Further, the obligatory and non-deliberate processing of unattended information may still allow the organism to learn, implicitly (i.e., without having to be consciously aware that it is processing the information), about statistical regularities in its environment, information that mediates the critical ability to predict and anticipate external events. Finally, such tension brings the critical capacity to have attention automatically or involuntarily captured by

events that may be more important than the currently selected information (e.g., it is crucial that a fire alarm could draw your attention away from the words on this page however absorbed you might be in this text!). A delicate balance must therefore be struck between two opposing requirements; the requirement for the integrity of attentional engagement on task-relevant information is set in opposition against the requirement that currently irrelevant information can compete for, and if necessary win, the control of action (Allport, 1989).

This action of the attentional system, one in which the paradoxical constraints of narrowness and breadth must somehow be simultaneously satisfied, is the context within which we consider the effects of noise on performance. The impact of preattentive processing is arguably greater in the auditory modality because we cannot easily prevent the sensory information entering the brain; we cannot easily shut or re-direct the ears to avoid sound in the same way as we can shut or move the eyes to avoid registering the reflection of light from a given object. Initially, our discussion draws extensively upon the results from the study of the so-called irrelevant sound effect (ISE; Colle and Welsh, 1976; Hughes and Jones, 2001; Jones, 1993, 1999; Salamé and Baddeley, 1982, 1989). Using this technique we have been able to substantiate the claim that unattended sound is organised into temporally-extended objects (or streams). In a subsequent section we then go on to describe in more detail the various ways in which the competition for the control of action by these irrelevant auditory objects brings indispensable benefits for cognitive functioning. Finally, we assert that while attentional selection to competition from irrelevant sound is an essential property of the cognitive system it comes at a price insofar as irrelevant sound can corrupt the integrity of certain important types of mental activity.

### **1. Decomposing sound: The formation of auditory streams**

For an organism's successful interaction with the external environment, its perceptual system must initially find solutions to the interrelated problems of 'what objects are out there?' 'where

are they?’ and ‘what, if anything, are they doing or having done to them?’ (i.e., what events are occurring?). The visual perceptual system has evolved to provide particularly rich solutions to the first two problems. Indeed, information about the spatial layout of and spatial interrelationships between objects (‘where?’), which in turn is used by the brain to decipher the form of individual objects and therefore eventually their identity (‘what?’) is topographically mapped onto the retinas and higher reaches of the brain (e.g., Bregman, 1990; Humphreys and Bruce, 1989). In other words, there is a great deal of spatial correspondence between the information projected onto the retinas and early visual brain areas and the disposition of objects in the world; only their relative distance is different.

For hearing, there is no such simple mapping of sensory space onto the receptor. Rather, partitioning the auditory world according to the spatially-distributed, sound-emitting objects within it is a matter of derivation from a compound signal (the pressure variations reaching the ear). But the process of derivation is more complex than that: objects, in the auditory modality, are not uniquely specified by space, as they are in vision. So, the successful partitioning of two auditory objects (say, two human voices coming from the same loudspeaker) that overlap in space, calls for some process of deriving the components; some computation and resolution of several parts must be achieved. Not only is the spatial correspondence different in sound and vision but the temporal domain is distinct also. In sum, visual objects, relatively speaking, tend to be ‘given’; the process of partitioning is relatively simple, but auditory objects require more derivation, more fissile processing, based on a host of interacting dimensions (pitch, intensity, timbre, space), that are dynamically time varying. In short, the perception of sound involves its partitioning into streams; discrete mental descriptions of each independent sound-emitting object in the environment (see Bregman, 1993, for an overview).

Most of the evidence for auditory stream formation has come from experiments in which participants are asked to report what they hear.

However, this approach does not reveal much about whether sound that is not currently being attended is also organised into coherent streams, or forms an undifferentiated mixture of sounds, although Bregman (1990) does speculate that the former is likely to be the case (see also Bregman and Rudnicki, 1975; but see Carlyon, Cusack, Foxton and Robertson, 2001, for a different view). We turn now therefore to discuss evidence garnered from our laboratory that provides compelling evidence that unattended sound is indeed organised into coherent streams.

### ***1.1 Evidence for preattentive stream formation: The irrelevant sound effect***

In order to assess the extent of preattentive processing generally, and more specifically whether or not sound is organised into discrete objects when unattended, the index of processing must be indirect, because asking participants about the unattended material clearly renders it no longer unattended (Jones, 1999). Our approach to this issue therefore has been to exploit the well established finding that irrelevant background sound that the participant is asked to ignore (and at no point is the participant asked anything about the sound) markedly disrupts performance of a visually-based serial recall task (Colle and Welsh, 1976, Jones, 1993, 1999; Salamé and Baddeley, 1982, 1989). The serial recall task involves presenting participants with seven to nine visual-verbal items (e.g., digits, letters or words) one by one on a computer screen (usually at the rate of about 1 per sec) which are to be recalled in strict serial order either immediately following the last item or sometimes following a short retention interval (typically 6 to 10 secs) during which participants are expected to keep subvocally rehearsing the sequence. This task has played a major role in cognitive psychological research because it is widely believed that the process of encoding and maintaining the order of events is critical to many if not most mental activities (e.g., the perception of speech and music, problem solving, reasoning, reading) as well as any motor activities in which the temporal sequencing of elements is crucial (e.g., speaking, writing, typing, driving; Baddeley, 1990; Crowder, 1976; Lashley, 1951; Henson, 1998). Not only are the

effects widely manifest, they are also marked in terms of the disruption they bring about. The disruptive impact of sound on performance is robust with the detriment reaching up to 30-50% compared to a quiet control condition (Ellermeier and Zimmer, 1997).

By now quite a substantial body of work has been undertaken using this technique. It seems reasonable to assert on the basis of this work that the key property of sound that endows it with disruptive potency is the presence of acoustical variation regardless of whether the sequence is made up of speech (e.g., *c, j, t, u* compared to *c, c, c, c*.) or non-speech tokens (a succession of tones changing in pitch versus a repeated tone; Jones, Madden and Miles, 1992). Other properties of sound such as the semantic and phonological properties of speech for example play little if any role in this effect (Buchner, Irmen and Erdfelder, 1996; Jones and Macken, 1995b; LeCompte and Shaibe, 1997; but see Hughes and Jones, 2003c). Notably, in terms of the implications of the research for applied settings, to which we return later, the intensity of the sound is not an influential factor (at least within the range of 48dB (A) to 76dB (A); Colle, 1980; Tremblay and Jones, 1999) and the effect is not evanescent; there is no evidence that the effect diminishes or 'habituates' with continued exposure to the sound (Hellbruck, Kuwano and Namba, 1996; Jones, Macken and Mosdell, 1997).

Most of the available data is best understood by supposing that there is a conflict between two seriation processes (those involved in retaining stimuli in serial order) and, somewhat counter-intuitively, the effect has nothing to do with the content of the individual tokens making up the relevant and irrelevant sequences (Hughes and Jones, 2003c; Jones, Beaman and Macken, 1996; Jones and Tremblay, 2000; see Hughes and Jones, 2001, for a critique of alternative theories). That is, hearing digits while trying to remember digits is no more disruptive to performance than hearing letters of the alphabet being spoken (but see Hughes and Jones, 2003c). Specifically, the changing-state account posits that the deliberate process of retaining the order

of the to-be-remembered items in the serial recall task (i.e., rote rehearsal) is corrupted by the obligatory and preattentive processing of the order of the changing tokens in the irrelevant sequence. Information pertaining to the order of irrelevant auditory tokens is assumed to arise as a by-product of integrating the tokens into a coherent stream despite the presence of some acoustical mismatch between successive tokens. In other words, change within the confines of a single auditory stream (i.e., changes on a common carrier) yields cues as to the order of the tokens within that stream (Jones, 1999).

### ***1.2 Evidence for the changing-state account***

Several lines of evidence converge to support the changing-state account. Two of these were considered in some detail in Hughes and Jones (2001) and so are only briefly noted again here. First, the changing-state effect increases as a function of the degree of acoustic difference between successive tokens in the irrelevant sound sequence, but only up to the point at which the difference becomes so large (in terms of location or pitch for example) that the successive items are likely to be partitioned into separate streams (Jones and Macken, 1995a, Jones, Alford, Bridges, Tremblay and Jones, 1999; Jones, St Aubin and Tremblay, 1999; Macken, Tremblay, Houghton, Nicholls and Jones, 2003). This modulation of the changing-state effect is explicable in terms of the impoverishment of information pertaining to the order of temporally successive events when those events traverse separate streams (Bregman, 1990). Second, only tasks that call upon or tend to encourage the adoption of a serial rehearsal strategy are sensitive to the changing-state ISE (Beaman and Jones, 1997, 1998; Jones and Macken, 1993; but see LeCompte, 1994). This is consistent with the view that the effect arises as the result of a conflict of two similar seriation processes. We now turn to consider two new lines of evidence that have served to buttress some of the key claims of the changing-state account.

#### ***1.2.1 Negative order-repetition priming: the legacy of preattentive coding***

Hughes and Jones (2003a) reasoned that if the order of irrelevant auditory items is encoded

preattentively, clearly a critical assumption in the changing-state account, then it may be possible to find evidence that such order information leaves a relatively long-lasting trace in memory. The approach taken was to use an ‘implicit repetition priming’ procedure in which the issue of whether unattended information has been encoded is tested by examining whether an individual’s ability to respond to some information differs as a function of whether or not that information has recently been presented as to-be-ignored (and thus unattended) information (see e.g., Driver and Tipper, 1989; Tipper, 1985; Tipper and Cranston, 1985). If a difference is found, then it can be inferred that the unattended material was indeed registered. In Hughes and Jones (2003a), it was found that the ability to recall in order a list of visual digits was poorer if that same list (the same digits in the same order) had been presented as an irrelevant auditory spoken sequence during the previous trial than if the to-be-remembered list contained the same digits as the previous unattended sequence but arranged in a different order. Thus, the specific order of the unattended auditory sequence must have been encoded for it to affect the recall of the same sequence when it was subsequently re-presented as a to-be-remembered list. Moreover, the fact that performance was worse rather than better in the repeated condition suggests that although the irrelevant order information is encoded, its representation is then immediately suppressed or inhibited so that the participant can focus more effectively on the task at hand (see Tipper, 2001). That is, the impairment to recall may have occurred because rehearsing the to-be-remembered list (which supports serial recall) is impeded due to the fact that the mental representation of that same order information was recently inhibited. Indeed, the process of inhibition may be essential in striking the delicate balance between the engageability and flexibility of attentional selection. That is, the competition for the control of action generated by the process of preattentively partitioning the world into objects is counteracted to some extent by an inhibitory process that acts on those objects that are currently irrelevant (Houghton and Tipper, 1994).

### ***1.2.2 Ease of identifying order within an attended sequence predicts its disruptive power as irrelevant sound.***

The ease with which order is identified in an attended sequence is modulated by the duration of its constituent elements. Specifically, participants presented with a repeated sequence of four vowels can readily report their order for vowel durations greater than 125 ms but performance is near chance for durations equal to or less than 100 ms (e.g., Thomas, Hill, Carroll and Garcia, 1970). Macken, Tremblay, Culling and Jones (2003) argued that if order information is encoded preattentively and it is this encoding that produces the changing-state ISE then the ease with which the order of a sequence of attended items can be identified should be reflected in the capacity of that sequence to disrupt recall when subsequently used as irrelevant sound. In Experiment 1 of Macken et al. the authors first replicated the vowel duration effect (e.g., Thomas et al., 1970), observing that order reports increased in accuracy as a function of vowel duration (75, 150, 300 and 500 ms) with a sharp improvement being seen in the 150 ms condition. With a new group of participants they then measured the disruptive potency of these four sequence-types when used as irrelevant sound in a serial recall task. It was found that the sequence for which order report was poorest (75 ms vowel duration condition) had no detrimental impact on serial recall when used as irrelevant sound whereas all other sequence-types gave rise to disruption.

Recent studies (Hughes and Jones, 2003a, c; Macken et al., 2003) provide compelling and arguably more direct evidence than was previously available that order information in auditory sequences can be encoded preattentively. The findings of Macken et al. (2003) also show more directly than the other lines of evidence that it is the availability of this order information that mediates an auditory sequence’s power to disrupt serial recall. Most importantly perhaps, Macken et al.’s procedure provides an independent index of order availability; the degree of order information in an auditory sequence can be measured via the order report task and the disruptive power of that

sequence as irrelevant sound can then be predicted from this independent index.

These recent strands of empirical evidence just discussed converge on the general conclusion that the order of irrelevant auditory elements is registered preattentively, and that this order information is a by-product of integrating the irrelevant elements into a coherent stream. In previous articles (e.g., Hughes and Jones, 2001; Jones, 1999; Macken, Tremblay, Alford, and Jones, 1999) we have tended to be preoccupied with emphasising the ‘negative side’ of this research, that is, how the preattentive organisation of irrelevant sound can corrupt short-term memory processing (and indeed we provide an overview of this aspect of the work again in Section 3). However, an obvious question which we have not addressed fully is ‘why would the perceptual system have evolved in such a way as to leave key mental processes open to corruption from irrelevant information?’ In the next section therefore we redress the balance by discussing three essential benefits that are brought by the preattentive organisation of sound into coherent objects.

## **2. Indispensable benefits of preattentively organising irrelevant sound**

### ***2.1 Flexible voluntary allocation of selective attention***

A critical capacity for flexible interaction with the environment is to be able to deftly re-allocate the focus of attentional selection at will (i.e., voluntarily) to other potentially interesting sources of information, that is, we must be able to disengage from attending to one object so that our actions, if we so wish, can then be based on other objects that were previously being ignored (Styles, 1997). A major debate in the selective attention literature generally has been concerned with the basis upon which the system can re-allocate the attentional focus, that is, how much information about currently unattended sensory inputs does the system require in order that selection can be re-allocated to pick out and select other, previously unselected, objects (Pashler, 1998; Styles, 1997)? The view that we promoted in Section 1, that the auditory scene is organised into coherent objects preattentively, is

supportive of the general model of selective attention in which selection of the object that is most relevant to current behavioural goals is based on a ‘choice’ of several coherently formed objects (e.g., Allport, 1989, 1993; Duncan, 1984; Houghton and Tipper, 1994; Humphreys and Bruce, 1989; Mondor and Terrio, 1998; Neumann, 1996; Vecera, 2000). In other words, the system has already organised the input into coherent objects before selection of the most behaviourally-relevant object is required.

This view contrasts with another broad class of attentional theories in which the most task-relevant object can be selected on the basis of a simple, rudimentary analysis of the environment into its physical features such as object location, colour (in vision), pitch (in audition), and so on (e.g., Broadbent, 1958; Treisman and Gelade, 1980). That is, on this view, such physical features need not be bound together into coherent objects before selection can take place. Evaluating the evidence for and against these two broad positions is beyond the scope of the present article (but see Allport, 1993; Pashler, 1998; Styles, 1997) but we would argue that the evidence and arguments for object-based selection are most compelling (see e.g., Houghton and Tipper, 1994; Mondor and Terrio, 1998; Vecera, 2000) and our own research on the processing of irrelevant sound, as discussed in section 1, also points to the conclusion that selection of auditory information (at least) is object-based. Indeed, it can be argued that if the currently unattended portion of the world was not represented as separate, coherent objects, it is difficult to see how attention could be re-allocated to those objects in the first place. Thus, whilst listening to a music concert for example we can, if we so wish, easily switch our attention between, say, the piano and the vocal soloist. This situation would also highlight how organising the currently unselected information still provides a functional ‘backdrop’ for the relevant sub-set of information; whilst focussing on the singer, the unattended piano (the ‘irrelevant’ object) is nevertheless complementing the melody conveyed by the vocalist (i.e., the sound from the piano does not simply ‘disappear’).

## 2.2 Sensitivity to transitional probabilities in 'irrelevant speech': Role in language acquisition

The success with which a given biological species (and individual organisms within that species) will survive and thrive is reliant on its capacity to familiarise itself with its surroundings so that it can effectively navigate through its environment and otherwise interact with the external world. One capacity that is likely to be crucial with respect to becoming attuned to the environment is the ability to learn about and exploit the statistical regularities of events that occur in that environment across both time and space (e.g., Shepard, 1981). In so doing, organisms, including humans, tend to settle into healthily stable and stereotypical behavioural patterns that are enslaved to the predictability of external events (e.g., Johnston and Hawley, 1994; Johnston and Strayer, 2001). Indeed, the process of auditory streaming may, to some extent at least, be based on learning the systematic way in which sounds that emanate from the same object or event tend to unfold, i.e., the auditory system becomes attuned to 'transitional probabilities'; knowledge about the likelihood (or probability) of a particular sound being heard given the nature of the previous one (i.e., a particular transition).

The capacity for auditory stream formation may in turn play an essential role in human language acquisition, clearly crucial for an infant's successful adaptation to its environment. Natural speech conveyed by a single speaker is characterised by certain sequential regularities or transitional probabilities in terms of its acoustic (timbre, pitch), phonotactic (e.g., Brent and Cartwright, 1996) and semantic structure (e.g., Levelt, 1989). Given the fact that fluent speech, unlike text, does not exhibit reliable spaces or pauses between words, a major problem the infant must solve is 'word segmentation', that is, the infant must somehow discern which sequences of sounds constitute words in its native language (Cole and Jakimic, 1980). One source of information available from continuous speech that infants as young as eight months old have been shown to exploit is that discrete sound segments (e.g., phonemes or syllables) that make

up words will co-occur with a greater probability than sound segments that traverse word boundaries. For example, if the infant hears a sequence such as 'pretty' followed on different occasions by various other words (e.g., 'pretty baby', 'pretty flower', 'pretty dress') it will have heard the co-occurrence of 'pre' and 'tty' (within-word sequence) relatively more often than the co-occurrence of, say, 'tty' and 'ba' or 'tty' and 'flo' and so on (across-word sequences; Harris, 1951; Jusczyk, 1999; Saffran, Newport, and Aslin, 1996). Information about such distributional cues may be a by-product of auditory streaming because the pattern of transitional probabilities within the spectral dimensions of sound across time is one of the main sources of information used by the perceptual system to partition the auditory scene into discrete streams (Bregman, 1990).

The experimental technique for demonstrating that infants can segment words on the basis of transitional probabilities in speech involves presenting them, in an 'exposure' phase of an experiment, with 'nonsense' words, e.g., *pabiku*, *tibudo*, *golatu*, *daropi*, played in a random order in a continuous loop of speech (i.e., with no pauses between the words) for two minutes. The triplets of syllables co-occurring most often within the speech stream would be those within the words, e.g., *pa-bi-ku*, whilst those across different words would occur relatively less often. That the infant can learn about this pattern of transitional probabilities is shown during a second 'test' phase of the experiment in which nonsense words are presented auditorily one by one, half of which are the words that were used in the speech stream (e.g., *pabiku*) and some of which are words composed of the last two syllables of one word and the first syllable of another, e.g., *bi-ku-ti*. By careful measurement of the infant's headturning behaviour towards each stimulus, which can be assumed to indicate its 'listening preferences', it has been shown that the infant successfully discriminates which are 'legal' words and which are not at an above chance level. In short, the infant is able to discern from a fluent stream of speech which sequences constituted discrete words based exclusively on transitional probabilities.

Most critical in the context of the present article is that such sensitivity to transitional probabilities has been shown to extend to a situation in which participants (this time children and adults were tested) are not attending to the speech during the ‘exposure’ phase but are engaged in a visual-based primary task (Saffran, Newport, Aslin, Tunick, and Barrueco, 1997). Thus it is possible that the presence of speech even when the infant is not deliberately listening to is nevertheless informing him or her about the structure of his or her native language.

### ***2.3 Involuntary allocation of selective attention: Attentional capture***

It has just been argued that it is functional for an organism to become attuned to the regularities in its environment across time and space so that it becomes adept at anticipating, processing and responding to environmental regularities. However, there must also be a mechanism in place that counteracts this bias toward the expected, normal, and commonplace, such that the attentional selection system can still be interrupted by unexpected, deviant and novel occurrences in the environment (Allport, 1989, Johnston and Strayer, 2001). The reason for this from an evolutionary perspective is that incoming stimulation that violates expectancies could often signal the presence of danger (e.g., a predator) or opportunity (e.g., prey) for the individual (Sokolov, 1963). Indeed, some authors have argued that this counteracting bias toward change detection acts at a global-evolutionary scale to prevent species/organisms from being over-entrenched in rigid maladaptive patterns of behaviour. That is, the balance that the healthy human cognitive system strikes between the capacity for attentional engagement and interruptibility (see e.g., Allport, 1989) may represent a microcosm of the balance that must be struck between stability and plasticity, between order and chaos (see Kauffman, 1993), at the level of the relationship between species and their ecosystems (Johnston and Strayer, 2001).

There is a very large literature documenting how unexpected changes in sensory stimulation can cause an involuntary shifting of the attentional

focus (i.e., attentional capture; for an overview, see Eimer, Nattkemper, Schroger, and Prinz, 1996) and we will not address it in detail here except to note that a recent study conducted in our laboratory provides some preliminary evidence for the general hypothesis that preattentive stream formation may act to constrain the attentional orienting response to change that signals the onset of an entirely new auditory object (Hughes and Jones, 2003b). The impetus for this study was the suggestion that if change detection can trigger an involuntary OR, could such a mechanism not underpin the changing-state irrelevant sound effect discussed in Section 1? That is, changing-state stimuli may continually capture attention away from the serial recall task and therefore produce the decrement in performance, whilst steady-state tokens would lead to habituation of the OR (a diminution of orienting with continued exposure to the same sound stimulus) thus performance would be spared (Cowan, 1995). However, there are several lines of evidence, which have been discussed in some detail elsewhere (e.g., Hughes and Jones, 2001), that refute this approach to the ISE. For example, a strong prediction of this account would seem to be that the impact of irrelevant sound stimuli should be attenuated or eliminated with pre-exposure to those stimuli or with repeated exposure to the same stimuli across an experimental session. This is based on the notion that repeated exposure to the same irrelevant stimuli should allow a mental description of the stimuli to be formed and retained hence stripping the stimuli of their novelty and therefore their power to capture attention. This prediction is not borne out by the experimental data however (Jones et al., 1997; Tremblay and Jones, 1998; but see Banbury and Berry, 1997).

Rather than being the result of attentional capture, the ISE seems to operate below the threshold of attentional capture by the irrelevant stimuli; as argued earlier, it reflects the cost of organising the irrelevant sound into objects without those objects necessarily capturing the focus of attention. Indeed, recall that a critical characteristic of the changing-state effect is that only change superimposed on a common carrier



(i.e., changes within the parameters of a single coherent stream) produces the effect (Jones and Macken, 1995a; Jones, Alford et al., 1999). Such changes, by definition, would not signal the onset of a new object; they are fluctuations within the spectral or temporal properties of a single object across time. Indeed, it would be dysfunctional for our attention to be drawn away involuntarily in response to every slight change in the environment.

We reasoned however that if only a change that cannot be accommodated within an extant stream is likely to capture attention then it should be possible to observe such an effect in the classical irrelevant sound paradigm if one of the irrelevant auditory tokens deviates in some way from the remainder of the irrelevant tokens. In support of this, Hughes and Jones (2003b) found that if a single irrelevant auditory item was presented out of time with the remainder of the irrelevant items, a marked cost to serial recall was observed over and above the basic changing-state effect. This effect is compatible with the suggestion that the violation of the otherwise regular temporal pattern formed by the irrelevant tokens was perceived as signalling the onset of a novel event thereby capturing attention from the primary task. It was also found that changing the global nature of the irrelevant sequence across trials also incurs a cost: If, having presented an irrelevant speech sequence in which the temporal gap between successive tokens was 500 ms for three successive serial recall trials, on the next trial the temporal gap between irrelevant items is changed to 125 ms, a cost to serial recall was incurred over and above the basic changing-state effect, as indexed by a subsequent recovery in performance with continued exposure to the 'new' sequence-type over the next two serial recall trials. Again, this latter finding is compatible with the notion that stimulation that cannot be accommodated within a template describing the parameters of a recently formed stream, thus presumably signalling the onset of a new auditory object, has the propensity to capture attention. The key implication of these findings then, is that organising unattended information into objects, which means that the

organism 'knows' what objects are currently out there, may serve to confine attentional orienting to changes in the environment that are likely to signal the onset of a new object.

### **3. Unavoidable costs of preattentively organising irrelevant sound: Corruption of short-term memory processing**

As described earlier, the claim that irrelevant sound is organised into auditory objects preattentively was based on research that has pin-pointed the properties of sound that do and do not endow it with power to disrupt performance of a serial recall task. In that section the focus was not on the disruptive aspect of the effects as such; the disruptive effect was used as a means of assessing the extent to which the sound was being perceptually organised despite being unattended. However, we also noted that the serial recall task is assumed to call upon a cognitive process that is crucial to many mental and motor activities, namely that of seriation. This means that whilst the process of preattentive auditory object formation may well bring indispensable benefits such as those discussed in the previous section, it also carries a cost when activities involving seriation must be performed in the presence of irrelevant sound. Given that we have recently discussed this aspect of the ISE in some detail in a recent article in this journal (Hughes and Jones, 2001) we only provide a brief overview of the main issues here.

#### ***3.1 The role of seriation and its corruption by irrelevant sound***

It has been argued that any task that involves having to keep a sequence of incoming stimuli in order (e.g., speech comprehension) or the temporal sequencing of motor plans (e.g., for speech production) will be vulnerable to the deleterious effect of preattentively seriating changing-state irrelevant sound. Essentially, it is thought that the process of seriating the task-relevant elements (e.g., individual digits or letters in the serial recall paradigm) involves imposing order onto elements whose transitional probabilities are low. These are elements that have not been perceived and/or produced in that order often enough to have a long-term memory representation of the entire sequence such as

would have occurred in the case of learning the alphabet, the days of the week, well-known sentences (e.g., 'Mary had a little lamb') and so on. Thus, when we encounter a telephone number we have not heard before, in order to remember it, we use habits and skills of speech to link the initially independent stimuli (the individual digits) into a coherent stream by articulating the number to ourselves repeatedly (rehearsal), and by grouping the number into two subsets perhaps, the aim being that recalling the number later will simply involve recalling one or two items or 'chunks' from long-term memory (see Frankish, 1989). Irrelevant changing-state sound disrupts this process because the perceptual system, beyond our control, is organising the changing items into a coherent stream, which involves seriating the irrelevant items. Insofar as this preattentive seriation process draws upon the same affordances as we rely upon to integrate the digits of a telephone number into a coherent unit for example, recall performance is compromised.

To the extent that seriation is involved to a larger or lesser degree in most mental tasks, we have argued that cognitive performance in many applied settings will be vulnerable to disruption by irrelevant changing-state sound (Hughes and Jones, 2001). These would include the open-plan office environment (see Banbury and Berry, 1997, 1998), call centres, the air traffic control tower (see Hughes and Jones, 2001), and the cockpit (Banbury and Jones, 2000).

### ***3.2 Independence of the ISE from intensity***

One of the first observations about the ISE that made it clear that it would call for an explanation in terms of how we process information as opposed to some general arousal mechanism was that the effect is independent of the intensity level of the irrelevant sequence (Colle, 1980; Tremblay and Jones, 1999). That is, above the absolute threshold of audibility, the organisation of irrelevant sound and its various beneficial and detrimental effects, seems to proceed regardless of the intensity of the sound. The early surge of scientific interest into the effects of noise on efficiency that was seen in the three decades following World War II was centred almost

exclusively on the effects of the loudness of the noise. Thus, the body of work we have described here (and recently in Hughes and Jones 2001) represents a shift of focus away from an emphasis on intensity (and constructs such as behavioural arousal; see Smith and Jones, 1992, for an overview of this early work) and toward explanations that emphasise the processing of information. Of course, the practical implications of the fact that these effects are not dependant on the sound being loud are immense, inasmuch as the sound has either to be below the threshold of audibility, or be masked in such a way as not to reveal the acoustic changes that are the basis of the 'changing state effect'. Whilst the former is expensive and complex, the latter may only serve to bring about losses in efficiency through the masking of sounds that may generally be irrelevant but on some occasions be relevant to the individual.

### ***3.3 Resistance of the ISE to habituation***

Another characteristic that has as much of a practical implication as a theoretical one, is that the disruptive effect of irrelevant changing-state sound is resistant to the process of habituation and therefore is not evanescent. That is, the disruptive effect will not diminish simply because an individual is exposed to the sound over extended periods. Habituation of the attentional orienting response no doubt plays a major role in mediating the engageability of attentional selection generally by preventing an organism from repeatedly orienting to information that, whilst originally novel and therefore potentially of importance, may have already been ascribed the status of irrelevant, uninteresting or unimportant (e.g., Näätänen, 1992). However, we have shown that there is nevertheless a residual cost associated with the presence of change within a single sound-emitting object (i.e., the 'changing-state effect') because such change, despite the fact it should not have the propensity to interrupt the organism, is nevertheless registered as a by-product of the essential process of preattentively generating *potential* candidate objects. This price is one worth paying because, as speculated earlier, one side-effect of preattentive object formation is that it means that attentional orienting can be

reserved for incoming information that signals the onset of a new object.

With regards to the practical implications of the foregoing, we would highlight the point that two potential sources of interference by irrelevant sound must be taken into account when considering noise abatement issues in applied settings. One source of interference comes from the preattentive registration of any acoustical change occurring within sound-emitting objects (the changing-state effect). This effect will not diminish with exposure to the sound-emitting object and will tend to selectively disrupt mental activities that call upon seriation processes. The second source of interference comes from the functional tendency to have one's attention captured by sudden *gross* changes in auditory stimulation even when they are irrelevant to a given individual, e.g., a colleagues' phone ringing following a period of quiet; a conversation starting up; a radio being turned on; workmen starting to drill out in the street, and so on. Note that this source of interference is likely to have a more general impact insofar as any mental activities, not merely those with a seriation component, should be interrupted by such changes. In this case however, if the stimulation continues, this type of interference will habituate. For example, once the onset of a conversation has captured our attention, our attention will tend to be captured by it less and less as that sound continues (because it soon loses its status as novel information). As noted above however, insofar as the sound of two voices will contain a great deal of changing-state information, the preattentive organisation of the voices will continue to exert a residual effect on the performance of any seriation-based mental task.

#### 4. Conclusions

From the standpoint of noise abatement, the study of auditory attention has illuminated a variety of issues. First, whether the precise mode of auditory interference occurs below or above the level at which the auditory stimulation captures our attention, auditory distraction is an inevitable side-effect of having a flexible and selective attentional system coupled to a sensory

modality (hearing) whose mode of processing is largely automatic. That is, the processing of sound appears to be obligatory; this has the undoubted advantage in that it is the first stage in the process of automatic perceptual organisation of unattended sound, which in turn permits the flexibility of attention. However, this advantage comes at the price of making the system more labile also. Further advantages to the obligatory processing of sound and its preattentive organisation are only now beginning to be recognised. Certainly there is evidence that we may learn from the statistical regularities of the auditory environment, as it were, without awareness. This is not to say we can learn facts necessarily, but we may be able to learn which sounds in our environment tend to co-occur, a capacity that may be crucially important to language development. In conclusion, our process-based approach to cognition seeks to set the distracting effect of noise in a broader context, one in which it can be viewed as a trade-off for the undoubted advantages conferred by the preattentive organisation of sound.

#### Author note

Robert W. Hughes and Dylan M. Jones, School of Psychology, Cardiff University, Wales, United Kingdom. Dylan Jones is also adjunct professor at the University of Western Australia. The research described here that was conducted at the Cardiff School of Psychology's Human Factors laboratory received financial support from the United Kingdom's Economic and Social Research Council. Correspondence can be addressed to Robert Hughes ([HughesRW@Cardiff.ac.uk](mailto:HughesRW@Cardiff.ac.uk)).

#### Correspondence Address

*R. W. Hughes*  
*School of Psychology, Cardiff University*  
*PO Box 901, Cardiff, CF10 3YG, UK*  
*Tel: (+44) 29 20874868;*  
*Fax: (+44) 29 20875858*

#### References

Allport, D. A. (1989). Visual attention. In M. I. Posner (Ed.), *Foundations of Cognitive Science*, (pp. 631-682). MIT Press.

- Allport, D. A. (1993). Attention and control: Have we been asking the wrong questions? A critical review of 25 years. In D. E. Meyer and S. Kornblum (Eds.), *Attention and performance, XIV*, 183-218. Cambridge, MA: MIT Press.
- Baddeley, A.D. (1990). *Human memory*. Hove, East Sussex: Erlbaum.
- Banbury, S., and Berry, D. C. (1997). Habituation and dishabituation to speech and office noise. *Journal of Experimental Psychology: Applied*, 3, 181-195.
- Banbury, S., and Berry, D. C. (1998). The disruption of speech and office-related tasks by speech and office noise. *British Journal of Psychology*, 89, 499-51.
- Banbury, S.P., and Jones, D.M. (2000). Driven to distraction. *Flight Deck International*, (April), 37-39.
- Beaman, C. P., and Jones, D. M. (1997). The role of serial order in the irrelevant speech effect: Tests of the changing state hypothesis. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 23, 459-471.
- Beaman, C. P., and Jones, D. M. (1998). Irrelevant sound disrupts order information in free as in serial recall. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, 51A, 615-636.
- Bregman, A. S. (1990). *Auditory scene analysis: The perceptual organisation of sound*. Cambridge, MA: MIT Press.
- Bregman, A. S. (1993). Auditory scene analysis: hearing in complex environments. In S. McAdams and E. Bigand (eds.), *Thinking in sound: The cognitive psychology of human audition*, (pp. 10-36), Oxford: Oxford University Press.
- Bregman, A. S., and Rudnicki, A. I. (1975). Auditory segregation: Stream or Streams? *Journal of Experimental Psychology: Human Perception and Performance*, 1, 263-267.
- Brent, M. R., and Cartwright, T. A. (1996). Distributional regularity and phonotactic constraints are useful for segmentation. *Cognition*, 61, 93-125.
- Broadbent, D.E. (1958). *Perception and communication*. Oxford: Pergamon.
- Buchner, A., Irmen, L., and Erdfelder, E. (1996). On the irrelevance of semantic information in the 'irrelevant speech' effect. *Quarterly Journal of Experimental Psychology*, 49A, 765-779.
- Carlyon, R.P., Cusack, R., Foxton, J.M., and Robertson, I.H. (2001). Effects of attention and unilateral neglect on auditory stream segregation. *Journal of Experimental Psychology: Human Perception and Performance*, 27, 115-127.
- Cole, R., and Jakimik, J. (1980). A model of speech perception. In R. Cole (Ed.), *Perception and production of fluent speech* (pp. 136-163), Hillsdale, NJ: Erlbaum.
- Colle, H. A. (1980). Auditory encoding in visual short-term recall: Effects of noise intensity and spatial location. *Journal of Verbal Learning and Verbal Behavior*, 19, 722-735.
- Colle, H. A., and Welsh, A. (1976). Acoustic masking in primary memory. *Journal of Verbal Learning and Verbal Behavior*, 15, 17-32.
- Cowan, N. (1995). *Attention and memory: An integrated framework*. Oxford: Oxford University Press.
- Crowder, R. G. (1976). *Principles of learning and memory*. Hillsdale, N.J.: Lawrence Erlbaum Associates Inc.
- Driver, J., and Tipper, S. P. (1989). On the nonselectivity of "selective" seeing: Contrasts between interference and priming in selective attention. *Journal of Experimental Psychology: Human Perception and Performance*, 15, 304-314.
- Duncan, J. (1984). Selective attention and the organization of visual information. *Journal of Experimental Psychology: General*, 113, 501-517.
- Eimer, M., Nattkemper, D., Schroger, E., and Prinz, W. (1996). Involuntary attention. In O. Neumann and A. F. Sanders (Eds.), *Handbook of perception and action (Vol. 3)*, (pp. 389-446). London: Academic Press.
- Ellermeier, W. and Zimmer, K. (1997). Individual differences in the susceptibility to the 'irrelevant speech effect'. *Journal of the Acoustical Society of America*, 102, 2191- 2199.
- Frankish, C. F. (1989). Perceptual organisation and precategorical acoustic storage. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 15, 469-479.
- Harris, Z. S., (1951). *Methods in structural linguistics*. Chicago: University of Chicago Press.
- Hellbruck, J., Kuwano, S., and Namba, S. (1996). Irrelevant background speech and human performance: Is there long-term habituation? *Journal of the Acoustical Society of Japan*, 17, 239-247.
- Henson, R. N. A. (1998). Short-term memory for serial order: The Start-End model. *Cognitive Psychology*, 36, 73-137.
- Houghton, G., and Tipper, S. P. (1994). A model of inhibitory mechanisms in selective attention. In A. Dagenbach and T. Carr (Eds.), *Inhibitory mechanisms in attention, memory, and language* (pp. 53-111). San Diego, CA: Academic Press.
- Hughes, R. W., and Jones, D. M. (2001). The intrusiveness of sound: Laboratory findings and their implications for noise abatement. *Noise and Health*, 13, 55-74.
- Hughes, R. W., and Jones, D. M. (2003a). A negative order-repetition priming effect: Inhibition of order in unattended auditory sequences? *Journal of Experimental Psychology: Human Perception and Performance*, 29, 199-218.

- Hughes, R. W., and Jones, D. M. (2003b). *Attentional capture by irrelevant sound: Violations of a stream-based neural model*. Manuscript in preparation.
- Hughes, R. W., and Jones, D. M. (2003c). *An order-Stroop effect: The impact of order incongruence between seen and heard sequences*. Manuscript submitted for publication.
- Humphreys, G. W., and Bruce, V. (1989). *Visual cognition*. Hove: Lawrence Erlbaum Associates.
- Johnston, W. A., and Hawley, K. J. (1994). Perceptual inhibition: the key that opens closed minds. *Psychonomic Bulletin and Review*, 1, 56-72.
- Johnston, W. A., and Strayer, D. L. (2001). A dynamic, evolutionary perspective on attentional capture. In C. Folk and B. Gibson (Eds.), *Attraction, Distraction, and Action: Multiple perspectives on attentional capture*, (pp. 375-397). Elsevier Science.
- Jones, D. M. (1993). Objects, streams and threads of auditory attention. In A.D. Baddeley and L. Weiskrantz (Eds.), *Attention: Selection, awareness and control*, (pp. 167-198). Oxford: Clarendon Press.
- Jones, D. M. (1999). The cognitive psychology of auditory distraction: The 1997 BPS Broadbent Lecture. *British Journal of Psychology*, 90, 167-187.
- Jones, D. M., Alford, D., Bridges, A., Tremblay, S., and Macken, W. J. (1999). Organizational factors in selective attention: The interplay of acoustic distinctiveness and auditory streaming in the irrelevant sound effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 25, 464-473.
- Jones, D. M., Beaman, C. P., and Macken, W. J. (1996). The object-oriented episodic record model. In S. Gathercole (Ed.), *Models of short-term memory* (pp. 209-238). London: Lawrence Erlbaum Associates.
- Jones, D. M., and Macken, W. J. (1993). Irrelevant tones produce an irrelevant speech effect: Implications for phonological coding in working memory. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 19, 369-381.
- Jones, D. M., and Macken, W. J. (1995a). Organizational factors in the effect of irrelevant speech: The role of spatial location and timing. *Memory and Cognition*, 21, 318-328.
- Jones, D. M., and Macken, W. J. (1995b). Phonological similarity in the irrelevant speech effect: Within- or between-stream similarity? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21, 103-115.
- Jones, D. M., Macken, W. J., and Mosdell, N. (1997). The role of habituation in the disruption of recall performance by irrelevant sound. *British Journal of Psychology*, 88, 549-564.
- Jones, D. M., Madden, C., and 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., Saint-Aubin, J., Tremblay, S. (1999). Modulation of the irrelevant sound effect by organizational factors: Further evidence from streaming by location. *Quarterly Journal of Experimental Psychology*, 52A, 545-554.
- Jones, D. M., and Tremblay, S. (2000). Interference by process or content? A reply to Neath (2000). *Psychonomic Bulletin and Review*, 7, 550-558.
- Jusczyk, P. W. (1999). How infants begin to extract words from speech. *Trends in Cognitive Sciences*, 3, 323-328.
- Lashley, K. S. (1951). The problem of serial order in behaviour. In L. A. Jeffres (ed.), *Cerebral mechanisms in behaviour: The Hixon Symposium* (pp. 112-146). New York: Wiley.
- LeCompte, D. C. (1994). Extending the irrelevant speech effect beyond serial recall. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 20, 1396-1408.
- LeCompte, D. C., and Shaibe, D. M. (1997). On the irrelevance of phonological similarity to the irrelevant speech effect. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, 50A, 100-118.
- Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. Cambridge, Mass., MIT Press.
- Macken, W. J., Tremblay, S., Alford, D., and Jones, D. M. (1999). Attentional selectivity in short-term memory: Similarity of process, not similarity of content, determines disruption. *International Journal of Psychology*, 34, 322-327.
- Macken, W. J., Tremblay, S., Culling, J., and Jones, D. M. (2003). *Disruption of order memory by task-irrelevant auditory sequences: A case of conflict of process*. Manuscript submitted for publication.
- Macken, W. J., Tremblay, S., Houghton, R. J., Nicholls, A. P. and Jones, D. M. (2003). Does auditory streaming require attention? Evidence from attentional selectivity in short-term memory. *Journal of Experimental Psychology: Human Perception and Performance*, 29, 43-51.
- Mondor, T. A., and Terrio, N. A. (1998). Mechanisms of perceptual organization and auditory selective attention: The role of pattern structure. *Journal of Experimental Psychology: Human Perception and Performance*, 24, 1628-1641.
- Näätänen, R. (1992). *Attention and brain function*. New Jersey: Lawrence Erlbaum Associates.
- Neumann, O. (1996). Theories of attention. In O. Neumann and A.F. Sanders (eds.) *Handbook of perception and action (Vol. 3)*, (pp. 389-446). London: Academic Press.
- Pashler, H.E. (1998). *The psychology of attention*. Cambridge, MA: MIT Press.

- Salamé, P., and Baddeley, A. D. (1982). Disruption of short-term memory by unattended speech: Implications for the structure of working memory. *Journal of Verbal Learning and Verbal Behavior*, 21, 150-164.
- Salamé, P., and Baddeley, A. D. (1989). Effects of background music on phonological short-term memory. *Quarterly Journal of Experimental Psychology*, 41A, 107-122.
- Saffran, J. R., Newport, E. L., and Aslin, R. N. (1996). Word segmentation: The role of distributional cues. *Journal of Memory and Language*, 35, 606-621.
- Saffran, J. R., Newport, E. L., and Aslin, R. N., Tunick, R. A., and Barrueco, S. (1997). Incidental language learning: Listening (and learning) out of the corner of your ear. *Psychological Science*, 8, 101-105.
- Shepard, R. N. (1981). Psychophysical complementarity. In M. Kubovy and J. R. Pomerantz (Eds.), *Perceptual organization*, (pp. 279-341). Erlbaum: Hillsdale, NJ.
- Smith, A. P., and Jones, D. M. (1992). Noise and performance. In A. P. Smith and D. M. Jones (Eds.), *Handbook of Human Performance* (Vol. 1, pp. 1-27). London: Academic Press.
- Sokolov, E. N. (1963). *Perception and the conditioned reflex*. London: Pergamon Press.
- Styles, E. (1997). *The psychology of attention*, Hove, UK: Psychology Press.
- Thomas, I. B., Hill, P. B., Carroll, F. S., and Garcia, B. (1970). Temporal order in the perception of vowels. *Journal of the Acoustical Society of America*, 4, 1010-1013.
- Tipper, S. P. (1985). The negative priming effect: Inhibitory priming by ignored objects. *Quarterly Journal of Experimental Psychology*, 37A, 571-590.
- Tipper, S. P. (2001). Does negative priming reflect inhibitory mechanisms? A review and integration of conflicting views. *Quarterly Journal of Experimental Psychology*, 54(A), 322-341.
- Tipper, S. P., and Cranston, M. (1985). Selective attention and priming: Inhibitory and facilitatory effects of ignored primes. *Quarterly Journal of Experimental Psychology*, 37(A), 591-611.
- Treisman, A. M., and Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology* 12, 97-136.
- Tremblay, S., and Jones, D. M. (1998). Role of habituation in the irrelevant sound effect: Evidence from the effects of token set size and rate of transition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 24, 659-671.
- Tremblay, S., and Jones, D. M. (1999). Change of intensity fails to produce an irrelevant sound effect: Implications for the representation of unattended sound. *Journal of Experimental Psychology: Human Perception and Performance*, 25, 1005-1015.
- Vecera, S. P. (2000). Toward a biased competition account of object-based segregation and attention. *Brain and Mind*, 1, 353-384.