Auditory attention

Beaman, P. ORCID: https://orcid.org/0000-0001-5124-242X (2021) Auditory attention. In: Braddick, O. (ed.) Oxford Research Encyclopedia of Psychology. Oxford University Press, Oxford. doi: https://doi.org/10.1093/acrefore/9780190236557.013.778 Available at https://centaur.reading.ac.uk/97021/

To link to this article DOI: http://dx.doi.org/10.1093/acrefore/9780190236557.013.778

Publisher: Oxford University Press

Auditory attention

C. Philip Beaman,

School of Psychology & Clinical Language Sciences, University of Reading, Reading UK

Summary

The modern world is noisy. Our streets are cacophonies of traffic noise, our homes and workplaces replete with bleeping timers, announcements, and alarms. Everywhere there is the sound of human speech – from the casual chatter of strangers and the unwanted intrusion from electronic devices through to the conversations with friends and loved ones we may actually wish to hear. Unlike vision, it is not possible simply to “close our ears” and shut out the auditory world and nor, in many cases, is it desirable. On the one hand, soft background music or environmental sounds, such as birdsong or the noise of waves against the beach, is often comfortingly pleasurable or reassuring. On the other, alarms are usually auditory for a reason. Nevertheless we somehow have to identify, from amongst the babble that surrounds us, the sounds and speech of interest and importance and to follow the thread of our chosen speaker in a crowded auditory environment. Additionally, irrelevant or unwanted chatter or other background noise should not hinder concentration on matters of greater interest or importance – students should ideally be able to study effectively despite noisy classrooms, or University halls, while still being open to the possibility of important interruptions from elsewhere. The scientific study of auditory attention has been driven by such practical problems: how we somehow manage to select the most interesting or most relevant speaker from the competing auditory demands made by the speech of others, or isolate the music of the band from the chatter of the nightclub. In parallel, the causes of auditory distraction – and how to try and avoid it where necessary – have also been subject to scrutiny. A complete theory of auditory attention must account for the mechanisms by which selective attention is achieved, the causes of auditory distraction, and the reasons why individuals might differ in their ability in both cases.

Keywords: Selective attention, auditory distraction, dichotic listening, irrelevant sound, informational masking, load theory of attention.

Introduction

The phenomenon of attention, and of “paying attention” was considered by James (1890) who wrote that, “My experience is what I agree to attend to. Only those items which I notice shape my mind — without selective interest, experience is an utter chaos”, indicating the importance he attributed to *selective* attention and further that "It implies withdrawal from some things in order to deal effectively with others, and is a condition which has a real opposite in the confused, dazed, scatter brained state...” which, similarly, implies an interest in *distraction*. As a subject of scientific study, however, attention only received serious experimental investigation following the so-called cognitive revolution of the 1950s (Shallice & Cooper, 2011). To “pay attention” to something as James described it, sounds at first blush like a relatively trivial instruction, implying as it does the need to prioritise the particular stimuli over others. Early research into this apparently commonplace activity rapidly ran into problems, however.

Much of the pioneering research on this topic took place in the 1950s and was driven in part by practical questions about multi-message environments encountered by fighter pilots. The research, conducted at the Applied Psychology Unit at Cambridge by Donald Broadbent, naturally focused upon the auditory, rather than the visual world (e.g., Broadbent, 1958; Cherry, 1953; Deutsch & Deutsch, 1963). Although latterly investigations into vision and visual attention have become more commonplace (e.g., Lavie, 1995) it is an open question whether the two sensory modalities follow similar principles in “attending to” one amongst many competing stimuli. Unlike vision, there is no eye-movement proxy for shifting auditory attention (at least, not amongst humans). Also, unlike vision, there is no simple way to cut-off auditory stimulation, leading to audition’s characterization as the sense most alert to early warnings or “the sentinel of the senses” (Handel, 1990).

For the researchers of the 1950s and 1960s concerned with practical problems of communication to pilots of fighter planes, the primary research goals were to identify on what basis a stimulus might be identified and selected for priority processing. From this followed the theoretical goal of determining what, precisely was the fate of the unattended (Broadbent, 1958). For example, if incoming stimuli are typically processed only in terms of their physical characteristics then something further is required to single out a stimulus for semantic analysis. This stimulus, which is analysed semantically, can then be said to be "attended" whereas other stimuli that are not subject to such analyses can be said to be "unattended". Although this early research generated much data on how and when section might occur, the topic of auditory distraction (arguably, a failure of selective attention) only attracted a similar level of systematic interest much later. More recent practical applications of such research have considered auditory distraction in everyday life such as the effects of mobile phones on driving (Horswill & McKenna, 1999; Strayer & Johnson, 2001) noise in office settings and classrooms (Banbury & Berry, 1998; Jahncke, Hygge, Halin, Green & Dimberg, 2011; Shield & Dockrell, 2003) as well as safety-critical situations, for example in the aviation industry (Hodgetts et al., 2005; Tremblay, Parmentier, Hodgetts, Hughes & Jones, 2012) and in health-care (Healey, Sevdalis & Vincent, 2006)

Auditory Selective Attention: Basic Findings

Broadly speaking, in humans, auditory information is processed in the primary auditory cortex, located in the temporal areas of the brain. Information is carried from the ears via connections to both left and right hemispheres with stronger connections contralaterally from the left ear to the right hemisphere (and from the right ear to the left hemisphere) than the ipsilateral connections between the ears and their corresponding cerebral hemispheres. The left hemisphere is usually (in about 95% of individuals depending on handedness; Corballis, Badzakova-Trajkov & Häberling, 2012) the hemisphere responsible for analysing speech for its lexical content and meaning, whereas the right hemisphere responds primarily to physical aspects of the auditory signal such as dynamic changes in frequency and pitch. Such dynamic physical characteristics also carry semantic information by means of prosody, which conveys emotion and can also alter the intended meaning of speech – the extent of which is dependent upon the language. This lateralisation of processing seems to hold regardless of whether the auditory information is a signal to be attended, or one which acts to "mask" the to-be-attended target (Scott, Rosen, Beaman, Davis & Wise, 2009). In the auditory domain it is relatively easy to mask a particular auditory stimulus by playing other sounds (e.g., white noise, competing speech) simultaneously. In studies of speech perception, sounds that mask the auditory events of interest at a relatively peripheral stage of processing in terms of the underlying physiology (e.g., at the level of the cochlea), and do so because two utterances contain energy in the same frequency bands, are referred to as providing "energetic masking". Sounds that mask more centrally, when both messages are audible but confusable are denoted "information masking" and the two are discriminable both behaviourally (Brungart, 2001) and neurally (Scott et al., 2009). Studies of auditory attention usually aim to minimise obviously perceptual problems of energetic masking, although the issue of whether informational masking can be equated with failure of selective or divided attention is more complex.

In a typical experiment of the 1950s or 1960s, using a procedure known as dichotic listening, participants would be asked to repeat aloud or "shadow" an auditory message presented to one ear (or "channel") over headphones while trying to ignore a masking or distracting message sent via a different channel (i.e., to the other ear). When a target and a masker are presented to both ears, target-signal detection is improved when there is a difference between the interaural phase of the target and the interaural phase of the noise, a phenomenon known as the binaural masking level difference (BMLD; Hirsh, 1948a,b; Licklider, 1948). Thus the dichotic listening procedure minimises energetic masking by means of spatially separating the two messages, although informational masking effects are clearly identifiable (Treisman, 1964). Typically, participants were unable to subsequently recall much, if anything, of the ignored or "unattended" message even if the content was repeated numerous times (Moray, 1959) and these results have proved to be very robust. For example, in replication studies carried out much later, and with improved technical control over experimental conditions, Wood and Cowan (1995a) found that about two thirds of people failed to notice that the speech stream in the unattended ear was reversed during the time they were shadowing speech from the attended ear. Participants were also unlikely to notice if the language of the unattended message switched from English to German, although they did notice if the speaker's gender changed, or if a tone or bleep was inserted (Cherry, 1953). This observation roughly accords with the famous failure in visual attention to notice a gorilla walking across a basketball pitch while counting the passes between members of a one of the basketball teams (Simons & Chabris, 1999; see also Rock & Guttman, 1981). A more exact translation of the gorilla experiment back into the auditory domain, in which participants fail to "hear" a male voice announcing "I am a gorilla" while focussing on one of two concurrent conversations, has also been reported (Dalton & Fraenkel, 2012).

Early Selective Attention and Filter Theory

To account for such findings, Broadbent (1958) proposed that auditory attention operated a system of *early* selection. His "filter theory" of selective attention, shown in Figure 1, is based upon three basic principles:

1) Stimuli presented concurrently gain access in parallel to a sensory buffer
2) There is a limited capacity to process stimuli beyond this point

3) To protect this limited capacity system, stimuli pass through a filter (or are selected) on the basis of their particular physical characteristics.

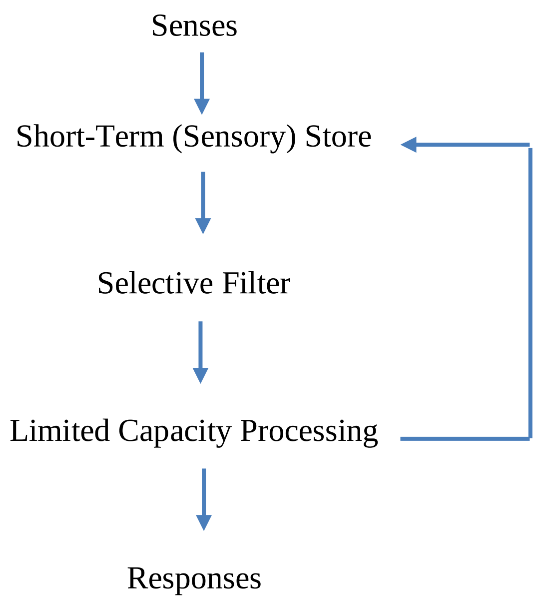

Figure 1. Visual depiction of Broadbent's filter theory of selective attention, credited as being the first "box-and-arrow" information-processing model.

The filter theory explains Cherry's findings that messages presented to the same ear can be separated out on the basis of physical cues such as sex of the speaker (i.e. voice pitch) or voice intensity (loudness), or spatial location (e.g., when presented to different ears in early experiments or via different loud-speakers in later studies), but not on the basis of meaning. It also explains why if one message is shadowed, little information is extracted or retained from the second message. However, the filter theory assumes that the unattended message is always rejected at an early stage in processing because the filter is a structural feature of the cognitive system. While Underwood (1974) reported that naïve subjects could detect only 8% of target digits presented in a non-shadowed message, Moray, a more practised shadower, could detect 67% of such digits. This implies, at the very least, a role for practise in the auditory selection tasks used to examine auditory attention. As already noted, Wood and Cowan (1995a) found that two thirds of individuals failed to detect when unattended speech was reversed, but this means that about a third of individuals did report noticing such changes. Wood and Cowan (1995b) similarly reported that 34.6% of their participants could recall hearing their own name on the unattended channel compared to 33.3% of Moray's (1959) participants.

Again, more recent data support this idea that attention might be more flexible. Studies of individual differences using the complex working memory span task presumed to index attentional control (Engle, 2002) have found both that participants with greater working memory capacity are less likely to notice their own name on an "unattended" ear while shadowing (Conway, Cowan & Bunting, 2001) and that they are more likely to report noticing such a target when they are asked to divide attention and try to monitor both channels (Colflesh & Conway, 2007). The former finding presumably reflects tighter control over the positioning or maintenance of auditory attention, while the latter suggests corresponding differences in the ability to divide attention.

One problem that was recognised early on with the conclusions from the dichotic listening procedure was that interpretation of the results is dependent upon participants' failure to overtly respond to, or recall, the content of the "unattended" stimuli. Logically, it is entirely possible that supposedly unattended stimuli are fully processed in terms of their meaning, but without conscious awareness. In an ingenious study of covert responses to the content of the unattended message Von Wright, Anderson & Stenman (1975) presented a long list of words for participants to study. When the Finnish word for "suitable" was presented, the participants received an electric shock, thus establishing a conditioned response to this particular utterance. In the second stage of the study participants were to shadow one word list, and to ignore another word list. When the Finnish word for "suitable" appeared in the to-be-ignored list the participants displayed a Galvanic skin response (GSR), thus showing that at some level, whether the individuals were conscious of this or not, the stimulus had been recognised, and its behavioural associations evoked. There are, however, problems with this and similar studies: Specifically, participants might occasionally shift attention to the "ignored" message (Cowan, 1995). It would require only a few occasions for this to occur to result in a difference between conditions in which no GSR was expected (e.g., when conditioned participants were presented with words other than the Finnish word for "suitable" – the conditioned stimulus –, or when non-conditioned participants were presented with the conditioned stimulus) and conditions in which a GSR might be seen either because of full processing of the unattended, or because of an attentional slip at that particular moment.

Of all the data that cast doubt upon a strict, early selection filer theory, however, the results obtained much earlier by Gray and Wedderburn (1960) are generally considered to be those that most effectively refuted filter theory. Gray and Wedderburn (1960) examined whether the meaning of two messages presented to separate ears could influence how they were organised. Participants were presented with messages to each ear and ask to repeat back what they could recall. For non-meaningful messages (i.e., lists of words which did not convey a coherent or meaningful sentences) participants typically reported words by ear – reporting first the words presented to the left ear and then to the right ear, for example, as would be predicted on the basis of early filter theory in which items are selected by physical features (such as spatial origin). If a meaningful message was separated across the two spatial locations (physical cues), such that a meaningful sentence was produced if participants alternated their reporting between ears (right, then left, then right again) – for example, "mice-3-eat-7-cheese-9" – the coherent meaning (mice eat cheese and 3 7 9) overrode the organisation by spatial location, as illustrated in Figure 2.

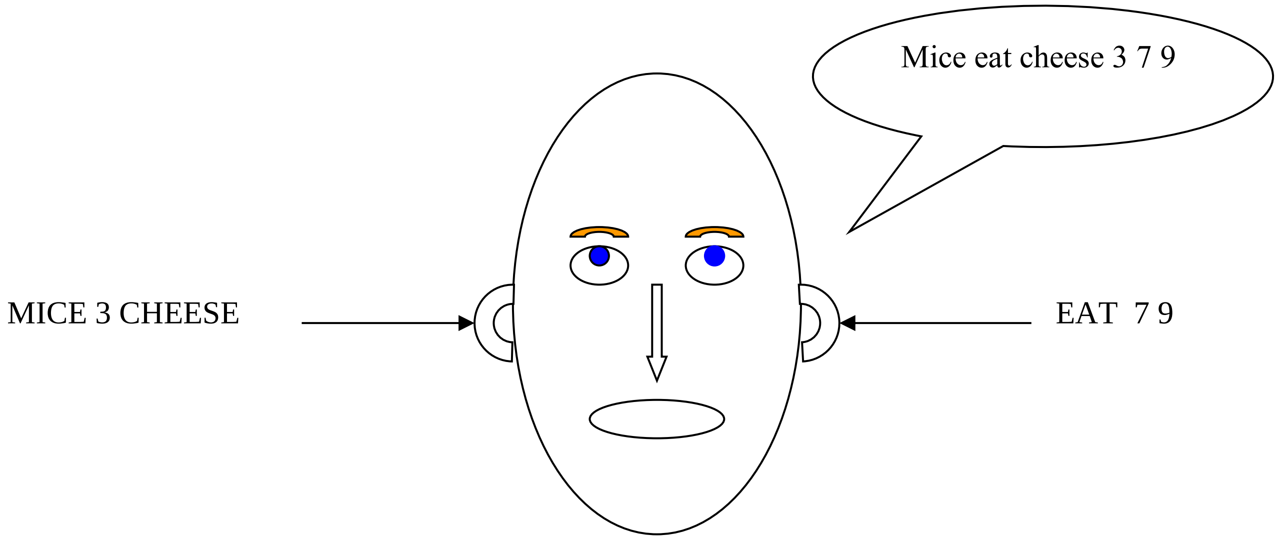

Figure 2. The participant is presented with words in both ears concurrently but groups their report semantically, so the meaningful messages “mice eat cheese” and “3 7 9” are reported rather than “mice 3 cheese eat 7 9” as would be reported if the messages were organised by spatial location (ear of presentation) rather than by meaning.

In other words, and contrary to the predictions of filter theory, the messages were selected according to meaning. In a related study, Treisman (1960) also found that if a meaningful message suddenly switched ears, participants in the experiment often switched ears in shadowing- contrary to instructions. Therefore, they must be selecting the message to attend to by meaning, not by physical cue. Again, this is not possible in Broadbent’s filter theory because in the filter theory selection occurs early, *before* analysis of the meaning of the message. This has left theorists with something of a quandary: The data that led to the formulation of filter theory, and early selection theories more generally, have proven to be robust, yet the theory itself has been manifestly disproved. Gray and Wedderburn’s (1960) data unequivocally point to selective attention operating at a late stage – beyond the point at which speech is analysed in terms of meaning.

In summary, early research provided evidence that experimental participants were consciously aware of the content of supposedly unattended auditory messages only in an (albeit substantial) minority of cases. This led to the formulation of the filter theory of early selective attention but this theory failed to account adequately for the minority of cases when participants could accurately report the content of the unattended and was also challenged by data showing that organisation of auditory messages by meaning, which *follows* organisation by physical features according to early selection theory, nevertheless took precedence in determining how individuals would recall information presented to them. Thus, although the filter theory inspired a tradition of information-processing models in psychology (Shallice & Cooper, 2011) the main contribution of research at this point was empirical data and an apparent theoretical puzzle.

Historically, two approaches were taken to resolving this seeming contradiction. One approach pioneered by Treisman (1960, 1964) was to amend, rather than reject, Broadbent's filter theory. Treisman (1960, 1964) added refinements to the theory such as attenuated processing, rather than wholesale rejection of information which did not pass through the filter. In her model, words are represented by dictionary units, with thresholds which must be reached for the unit to be activated. These thresholds are adjustable, such that one is more likely to become aware of important stimuli (e.g., one's own name) or sequentially very likely words in continuous speech even if presented as part of an unattended message. Thus, certain pertinent or expected stimuli may be noticed despite the attenuation of the incoming signal (see Figure 3).

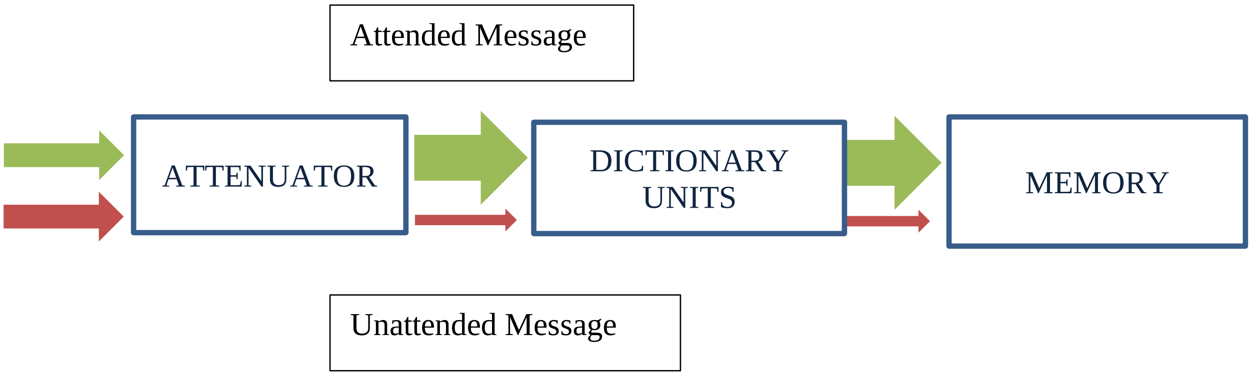

Figure 3. Treisman's attenuated filter theory. The figure shows the processing of two simultaneously presented messages – an attended message (above, in green) and an unattended message (below, in red). Subsequent to attenuation, the strength of the attended message is very much greater than that of the unattended message although dictionary units with sufficiently low thresholds might respond to signals from either source.

A second approach involves a more wholesale rejection of Broadbent's idea of a selective sensory filter. Instead, this approach suggests that all stimuli are fully processed but only the most important signal is selected for a response (Deutsch & Deutsch, 1963). In signal detection terms this can be understood as the difference between adjusting the sensitivity of a signal (which is maximised by an early or stimulus selection theory) and adjusting the decision criteria, or willingness to respond.

Load Theory and its Application to Audition

An ingenious approach to resolving this early versus late selection debate was proposed by Lavie (1995; Lavie & Tsal, 1994), originally in the context of visual selective attention. In her "load theory" of selective attention, she argued that there is obligatory processing of "irrelevant" stimuli provided there is some attentional resource available to carry out the processing. Only when attention is already overloaded will irrelevant stimuli be filtered or selected out. Thus, the idea is that rather than a strict filter as proposed by Broadbent (1958), or even an attenuating filter as suggested by Treisman (1960), there is more of a flexible gating system where the gate remains open – and all stimuli are fully processed – up until such point as too many stimuli are competing to gain access, in which case stimuli are subject to prioritisation and only those identified as requiring attention are allowed through. Hence, when attention is overloaded (as in many dichotic listening studies) it should appear that selection happens early. When attentional resources are sufficient to analyse messages fully (as, arguably, is the case in the Gray and Wedderburn (1960) experiment) then selection should appear "late".

Results from an early study by Treisman & Geffen (1967) to test whether selection occurred at perception (as in early selection) or at response (as in late selection theory), provide results explicable in terms of load theory. In their study Treisman & Geffen (1967) required subjects to shadow one of two messages, but to tap whenever they heard a target word in either message. According to late selection theory there is complete perceptual analysis of all inputs so there should be no difference in rates of detection between the shadowed and non-shadowed messages. According to load theory, there should be less opportunity to analyse the non-shadowed message and so fewer target words should be responded to in the non-shadowed message. In fact, detection rate in the shadowed message was 87%, in the non-shadowed message it was only 8%, results inconsistent with a later response selection but explicable within the framework of load theory.

More contemporary applications of load theory have repeatedly confirmed its utility for explaining visual attention (Murphy, Groeger & Greene, 2016). Its application to purely *auditory* attention and the resolution of the early-late selection debate in the auditory modality (in which the debate first arose) is more problematic. In the first place, defining an auditory perceptual load is less straightforward than for vision. Auditory scene analysis (Bregman, 1990) segregates the auditory environment into perceptual “streams” which allows for the identification of the origin and progression of a series of auditory events (or stream). Auditory scene analysis allows individuals to stream individual voices and hence to analyse each contribution to a conversation with minimal interference between the speakers but within each stream there are of course multiple auditory events (e.g., speech utterances). Although each event obviously contributes in some way to auditory perceptual load, the boundaries between each event are not always clear. This is readily apparent from any visual depiction of the speech signal – although the words are perceived as discrete utterances, the acoustic signal is in fact continuous over time as illustrated in Figure 4. Thus, although auditory perception can be cast in terms of the identification of auditory objects, analogous to visual objects, and there are clear commonalities between auditory and visual attention (Lee, Larson, Maddox & Shinn-Cunningham 2014; Shinn-Cunningham, 2008) the perceptual load both within and between streams is not so easily or unambiguously quantifiable. One way around this, discussed in the section on Distracted Attention, is to examine the effects of increasing visual perceptual load on auditory attention.

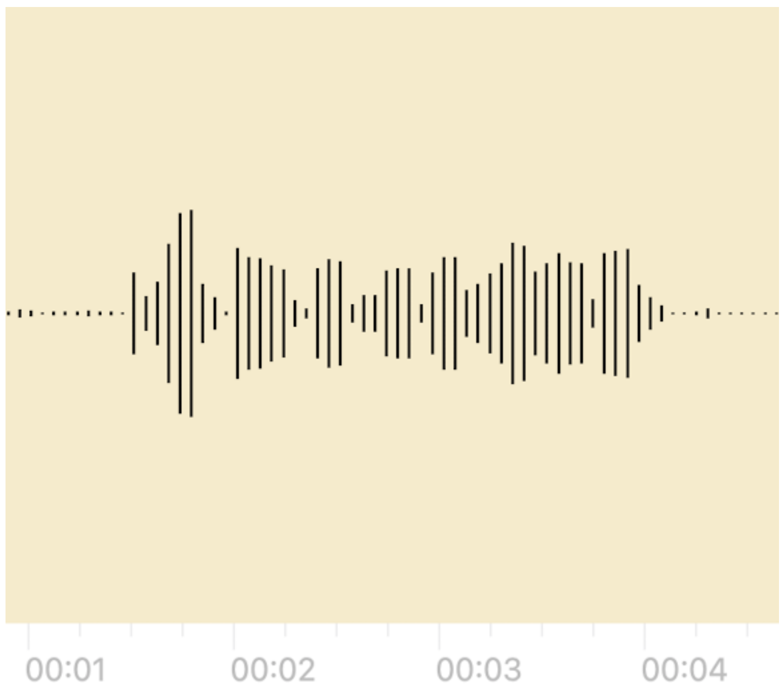

Figure 4. The waveform of the phrase "*The Quick Brown Fox Jumped Over The Lazy Old Dog*" spoken over a 4 second period by a male speaker in a UK English accent. Notice that although there are ten discrete words in this sentence, it is impossible to unambiguously identify ten discrete utterances from visual inspection of the waveform alone.

Secondly, attempts to directly test perceptual load theory with auditory stimuli have produced mixed results. Notably, Murphy, Fraenkel and Dalton (2013) failed to find, across four separate tasks, any modulation of auditory distractor processing as a function of perceptual load. These authors argued that, although it is possible to selectively process a preferred stream, the function of audition as an early-warning system or a "sentinel of the senses" implies that some processing capacity must always be somehow available in reserve regardless of perceptual load. This conclusion is somewhat at odds with the existing data showing how limited awareness sometimes is of unattended auditory information (e.g., Dalton & Fraenkel, 2012) and does not shed much light on the circumstances under which one might expect auditory information to go unnoticed. However, as already noted, load theory (and, more broadly, the early vs late selection debate) does not apply only within sensory modalities. In principle a high visual load might affect auditory selective attention or vice versa. Thus, Molloy, Griffiths, Chait & Lavie (2015) found that when the perceptual load in a visual search task was high there was a failure to perceive task-irrelevant tones, which was accompanied by a reduction in auditory evoked potentials, electrical activity within the brain evoked by the presence of auditory stimuli and measured using magnetoencephalography (MEG).

Distracted Attention: The Effects of Irrelevant Sound

An alternative means of looking at the extent to which auditory stimuli that are to be ignored are, nevertheless, subject to processing is by looking at the consequences that James (1890) identified as following from a failure of selective attention: auditory distraction. Auditory distraction operates cross-modally, allowing for manipulations of the difficulty of a visual task (e.g., visual perceptual load) and examining how this interacts with auditory distraction effects. In a typical auditory distraction study, a visually-presented and attention-demanding memory task is presented for participants to attempt and, alongside this, speech or other auditory stimuli are played. Participants are asked to try and ignore anything they may hear and reassured they will not be tested on it in any way (e.g., Jones & Macken, 1993). The extent to which the presence of speech or other auditory events interferes with the ongoing visual processing task is taken as an indication of the extent to which auditory distractors are subject to processing. This procedure ensures that energetic masking of the to-be-remembered targets by the distractors is not an issue because the primary task is visual, but it shares two limitations of the dichotic listening task.

The first limitation is that, as with dichotic listening, it is only apparent that the unattended stimuli are processed in some way when they have some measurable effect – in this case distraction. A lack of any distraction effect does not however, necessarily, imply a lack of processing. For example, it is clear that meaningful stimuli are more distracting primarily when they compete with other semantic analyses, as shown in the effects of the meaning of unattended speech on proof-reading tasks (e.g., Jones, Miles & Page, 1990) that are not always evident on other tasks less reliant upon semantics, such as immediate serial recall (e.g., Buchner, Irmen & Erdfelder, 1996). One possible criticism of this interpretation is that the apparent effect of meaning in the auditory distraction of proof-reading might be specific either to the distractors chosen or to the relationship between the distractors and the visual text. This concern is allayed, however, by studies showing that meaningful distractors are also more disruptive in a verbal free recall task where the meaning of the to-be-remembered targets is emphasised than when the same stimuli are used but recall of the order in which the targets were presented is emphasised (Marsh, Hughes & Jones, 2008).

The second limitation is simply that (as with Cowan's (1995) criticism of some dichotic listening tasks) the extent to which the to-be-ignored, irrelevant sound is actually unattended is not clear. This is reflected in the evolving terminology – from referring to auditory distraction as the "unattended speech effect" (Colle, 1976; Colle & Welsh, 1980; Salamé & Baddeley, 1982) to the "irrelevant speech effect" (Jones & Morris, 1992) in recognition that the attended (or unattended) nature of the speech should not be prejudged. The issue of determining what is, and what is not, unattended is not limited to auditory attention (e.g., Driver & Tipper, 1989) and most recently the "irrelevant sound effect" (Beaman & Jones, 1997) has gained prominence, in recognition of the fact that that not all auditory distraction is speech. This final shift in terminology also highlights an advantage of this procedure in that it can be (and has been) used to examine the processing of all kinds of non-speech auditory stimuli, in a way that has not typically been the case with dichotic listening. This approach also has the advantage that participants are never queried about the unattended stimuli so that any effects that are observed are not reliant upon conscious awareness or subsequent recollection of the unattended (Beaman & Jones, 1997).

Although the caveats raised above need to be borne in mind, studies of auditory distraction have nevertheless revealed several significant features of how to-be-ignored auditory stimuli interfere with ongoing cognitive processing. Speech is neither necessary nor sufficient to engender distraction (Jones & Macken, 1993) but the kind of dynamic and abrupt pitch changes common in speech create significant disruption (Jones, Macken & Murray, 1993). For example, continuous pitch glides are significantly less distracting than the same pitch glides when interrupted by periods of silence (Jones et al., 1993). Speech played to the right ear is also less disruptive to verbal short-term memory than speech played to the left (Hadlington, Bridges & Darby, 2004). Given that many aspects of verbal processing are the responsibility of the left hemisphere and that – as with vision – connections between the ears and the cerebral hemispheres are principally contralateral rather than ipsilateral, this finding that speech played to the left ear is more rather than less distracting is rather surprising. It becomes less so when one recalls that dynamic pitch changes of the kind which cause such distraction (Hadlington, Bridges & Beaman, 2006) are registered primarily in the right hemisphere (Scott et al., 2008).

One hypothesis put forward on this basis was that – since the left hemisphere is largely responsible for lexical processing – any distraction which has its locus in the processing of lexicality or meaning should, in contrast to the basic acoustic effect, be more pronounced if the to-be-ignored stimuli are played to the right ear (Beaman, Bridges & Scott, 2007). This prediction was subsequently independently confirmed (Sörqvist, Marsh & Jahnke, 2010).

Studies of auditory distraction have also confirmed the key role in shaping auditory attention played by auditory streaming, or the process by which the auditory system organises sound into elements that group or "stream" together (Bregman 1990). Whether streaming can occur in the complete absence of attention remains controversial (see, for example, the debate between Macken et al. (2003) and Carlyon, Cusack, Foxton and Robertson (2001)) but streaming of to-be-ignored voices clearly occurs, as shown by the additional disruption caused to an ongoing task by the addition of extra voices (Jones & Macken, 1995). The additional disruption is observed provided that each voice is associated with a discrete spatial location. When each extra voice is played as apparently coming from the same location, distraction is actually reduced – presumably because the perceptual "babble" which results smooths the acoustic signal and results in fewer perceptible changes within the stream. In contrast, where each additional voice is associated with a different stream multiple changes are perceived per stream (Jones & Macken, 1995). Additionally, where a superficially cohesive auditory stream (e.g., from a single spatial location) breaks down because the components of that stream clearly cannot originate from a single source (e.g., they differ wildly in terms of timbre) then this also modulates the disruptive effect of the irrelevant sound (Jones, Alford, Bridges, Tremblay & Macken, 1999). With respect to load theory, various manipulations of load in a visually-presented task reduce some forms of auditory distraction but not others (Hughes et al., 2013; Hughes, 2014; Marsh et al., 2018, 2020), going against load theory's assumption that high load *necessarily* acts to filter out the processing of task-irrelevant sound in all cases.

Although the irrelevant sound effect is reliable at the group-level, as with dichotic listening, there are also reliable, and well-documented individual differences in distractibility (Ellermeier & Zimmer, 1997). Unlike the ability to focus and share attention in dichotic listening tasks (Conway et al., 2001; Colflesh & Conway, 2007) these differences are not, however, associated with differences in working memory capacity for the most part (Beaman, 2004) and the basis for these differences remains largely mysterious, although there are undoubtedly strategic reactions to try and minimise any perceived distractions and these may vary between individuals (Hanczakowski, Beaman & Jones, 2014; 2018).

Training on a dichotic-listening type selective attention task has also recently been shown to attenuate the irrelevant sound effect with speech stimuli relative to the effect observed with an active control condition (Kattner & Ellermeier, 2020), suggesting that there is some learnable, top-down capability to "tune out" at least some forms of unwanted sound even if the bases for individual differences in the ability to do so remain to be discovered. It has also been suggested that auditory distraction may take multiple forms depending upon the nature of the distractor. Beaman (2004) noted that although working memory capacity did not mediate the irrelevant sound disruption effect in his studies, when the auditory distractors were semantically related to the to-be-recalled target material then participants with low working memory mistakenly reported the distractors as part of their recall protocol to a significantly greater extent than high working memory participants.

On the basis of findings such as these, a duplex account of auditory distraction has been put forward (Hughes, 2014), in which the effects of dynamic changes are distinct from the effects of simple predictability associated with an auditory "oddball" (Parmentier, Elsley, Andres & Barcelo, 2011). In the auditory oddball procedure, reaction times to a visually-presented target and event-related potentials recorded directly from the scalp both respond to an unpredictable or oddball event within an auditory stream. According to this theory, there is an irrelevant sound distraction effect associated with dynamic changes in the auditory environment but – overlaid upon this – there is at least one more effect that has to do with predictability rather than change per se.

This effect of unpredictability is observed alongside the distracting effects of change in irrelevant sound studies, where a predictably changing sound sequence continued to disrupt performance on a primary visual memory task but the insertion of an unpredictable auditory event added further distraction (Hughes, Vachon & Jones, 2007). Forcing participants to engage more fully with a primary task by varying the 21erceptibility of to-be-attended visual stimuli reduces the distraction associated with unpredictable auditory oddballs but not the distraction associated with predictable changes in the auditory stream (Hughes et al., 2013).

These findings suggest that the distraction associated by an unpredictable novel event is subject to a level of top-down control not observed with change *per se*, and align with suggestions from visual attention that selective attention should distinguish between bottom-up or stimulus-driven selection, which is largely what has been considered in traditional dichotic-listening tasks, and top-down or goal-driven selection (Lee et al., 2014). Hughes and colleagues (Hughes, Vachon & Jones, 2005; Vachon, Hughes & Jones, 2012) suggest that a version of Näätänen's (1990) neural model hypothesis can account for these findings. According to Näätänen's account, a neural model is constructed of the auditory environment and responses to mismatches between the neural model and the auditory input provoke an orienting response, which is habituated when there is a good degree of agreement between the incoming auditory stimuli and what is expected on the basis of the neural model. The duplex account of auditory distraction (Hughes, 2014) suggests that this aspect of distraction is open to top-down control (i.e., orienting responses can be inhibited if task demands require it) but that change *per se* in the auditory stream is unavoidably processed, and this may have an effect on concurrent processing.

Implications for the Unattended Message.

Substantial differences exist between the early dichotic listening experiments and the more recent focus on distraction by irrelevant sound, but nonetheless there are some common themes which emerge. When presented with multiple messages and asked to shadow one such message it is abundantly clear that individuals are unable reliably to identify changes in the meaning of a simultaneously presented but presumably unattended message, although it is much easier to notice physical changes in the unattended message. Similarly, in the majority of cases, individuals fail to notice the insertion of a significant target (the individual's own name) in the unattended message. Distraction studies show, however, that some processing of to-be-ignored auditory information takes place nevertheless. When tested for appropriately the presence even of ignored sound, and especially speech, interferes with ongoing processing of visually-presented information almost universally. Individuals also show good awareness of the distraction, as measured directly by self-report (Ellermeier & Zimmer, 1997), and indirectly by their willingness to volunteer answers to questions about visually-presented material and their confidence in the answers (Hanczakowski et al., 2014). Despite this, individuals may not always attempt to compensate for the distraction appropriately (Hanczakowski et al., 2018).

Results from studies of auditory distraction are also consistent with studies such as that of Gray and Wedderburn (1960). Gray and Wedderburn (1960) showed that people make use of semantic information to group information for recall, seemingly at odds with many earlier studies of selective attention. These early studies favoured an early-filter account based on the physical characteristics of the auditory signal, but many such studies also required participants to shadow a target message as an essential component of their experimental design. Shadowing the target message requires translating from auditory input into a motor (articulatory) output but, in such a task, the meaning of the message is less important than the ability to simply translate from the auditory input in the target ear to the articulatory configurations necessary to shadow the message. It is perhaps not surprising then if people lock onto and focus all of their attention on the physical cues necessary to conduct the focal (shadowing) task as a form of goal-driven, rather than stimulus-driven, selective attention. Under such circumstances, failure to respond to or recall the semantics of the distracting speech might reflect not so much the lack of any semantic processing as its irrelevance for the task in hand. The Gray and Wedderburn (1960) study did not involve shadowing but instead required participants to recall from a string of words simultaneously presented to both ears, with the assumption that selection for recall would follow the same principles as selection for shadowing. In many cases this might be true, but the ability to chunk items into larger meaningful units greatly aids recall (Miller, 1956) and therefore lends itself to obvious utilisation in participants' responses to this memory task.

Analogously, when asked recall a series of visually-presented words in a standard auditory distraction experiment there is usually no effect of the meaning of the distractor speech (Buchner et al., 1996) but for reasons already discussed this does not imply that the distractor speech is not processed semantically. The same distractors that show no effect of meaning when recall is based around simple, verbatim rote recall of visually-presented sequences show an additional disruptive effect of semantics if the recall task is switched to one in which memory for the meaning of the visually-presented sequences is required. Thus, what is processed – or more accurately, what is used to distinguish the attended from the unattended information – is dependent upon the task requirements, not just the stimulus characteristics as has been assumed. The "fate of the unattended" (Jones, 1995) is determined by the processing requirements of the focal task and the attended information.

References

Banbury, S. P., & Berry, D. C. (1998). Disruption of office-related tasks by speech and office noise *British Journal of Psychology, 89*, 499-517 doi.org/10.1111/j.2044-8295.1998.tb02699.x

Beaman, C. P. (2004). The irrelevant sound phenomenon revisited: What role for working memory capacity? *Journal of Experimental Psychology: Learning, Memory & Cognition, 30,* 1106-1118 doi.org/10.1037/0278-7393.30.5.1106

Beaman, C. P., Bridges, A. M., & Scott, S. K. (2007). From dichotic listening to the irrelevant sound effect: A behavioural and neuroimaging analysis of the processing of unattended speech. *Cortex, 43*, 124-134. doi.org/10.1016/S0010-9452(08)70450-7

Beaman C. P., & Jones, D. M. (1997). Role of serial order in the irrelevant speech effect: Tests of the changing-state hypothesis. *Journal of Experimental Psychology: Learning, Memory & Cognition*, 23, 459-471. doi.org/10.1037/0278-7393.23.2.459

Bregman, A. S. (1990). *Auditory scene analysis*. Cambridge, Ma.: MIT press.

Broadbent, D. (1958). *Perception and Communication.* London: Pergamon Press.

Brungart, D. S. (2001). Informational and energetic masking effects in the perception of two simultaneous talkers. *Journal of the Acoustical Society of America, 109,* 1101-1109. doi.org/10.1121/1.1345696

Buchner, A., Irmen, L., & Erdfelder, E. (1996). On the irrelevance of semantic information for the "irrelevant speech" effect. *Quarterly Journal of Experimental Psychology, 49A*, 767-779 doi.org/10.1080/713755633

Carlyon, R. P., Cusack, R., Foxton, J. M., & Robertson, I. H. (2001). Effects of attention and unilateral neglect on auditory stream segregation. *Journal of Experimental Psychology: Human Perception & Performance*, *27,* 115-127. doi.org/10.1037/0096-1523.27.1.115

Cherry, E. C. (1953). Some experiments on the recognition of speech with one and with two ears. *Journal of the Acoustical Society of America*, 25, 975–979. doi.org/10.1121/1.1907229

Colflesh, G. J., & Conway, A. R. (2007). Individual differences in working memory capacity and divided attention in dichotic listening. *Psychonomic Bulletin & Review,14*, 699-703. doi: 10.3758/bf03196824

Colle, H. A. (1980). Auditory encoding in visual short-term recall: Effects of noise intensity and spatial location. *Journal of Verbal Learning & Verbal Behavior, 19*, 722-735 doi.org/10.1016/S0022-5371(80)90403-X

Colle, H. A., & Welsh, A. (1976). Acoustic masking in primary memory. *Journal of Verbal Learning & Verbal Behavior, 15*, 17–31. doi.org/10.1016/S0022-5371(76)90003-7

Conway, A. R., Cowan, N., & Bunting, M. F. (2001). The cocktail party phenomenon revisited: The importance of working memory capacity. *Psychonomic Bulletin & Review, 8,* 331-335. doi.org/10.3758/BF03196169

Corballis, M. C., Badzakova-Trajkov, G., & Häberling, I. S. (2012). Right hand, left brain: genetic and evolutionary bases of cerebral asymmetries for language and manual action. *WIRES Cognitive Science, 3,* 1–17. doi: 10.1002/wcs.158

Cowan, N. (1995). *Attention and memory: An integrated framework.* Oxford: Oxford University Press.

Dalton, P & Fraenkel, N. (2012). Gorillas we have missed. Sustained inattentional deafness for dynamic events. *Cognition, 124,* 345-355. /doi.org/10.1016/j.cognition.2012.05.012

Deutsch, J. A., & Deutsch, D. (1963). Attention: Some theoretical considerations. *Psychological Review, 70,* 80–90 doi.org/10.1037/h0039515

Driver, J., & 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. https://doi.org/10.1037/0096-1523.15.2.304

Ellermeier, W., & Zimmer, K. (1997). Individual differences in susceptibility to the "irrelevant speech effect" *Journal of the Acoustical Society of America, 102,* 2191-2199. doi.org/10.1121/1.419596

Engle, R. W. (2002). Working memory capacity as executive attention. *Current Directions in Psychological Science, 11,* 19-23. doi.org/10.1111/1467-8721.00160

Gray, J. A., & Wedderburn, A. A. I. (1960). Grouping strategies with simultaneous stimuli. *Quarterly Journal of Experimental Psychology, 12,* 180-184.

Hadlington, L.J., Bridges, A., & Beaman, C. P. (2006) A left-ear disadvantage for the presentation of irrelevant sound: Manipulations of task requirements and changing state. *Brain & Cognition, 61,* 159-17 doi.org/10.1016/j.bandc.2005.11.006

Hanczakowski, M., Beaman, C. P., & Jones, D. M (2014) The effects of distraction on metacognition and metacognition on distraction: evidence from recognition memory. *Frontiers in Psychology, 5*. 439. doi: https://doi.org/10.3389/fpsyg.2014.00439

Hanczakowski, M., Beaman, C. P., & Jones, D. M. (2018) Learning through clamor: the allocation and perception of study time in noise. *Journal of Experimental Psychology: General, 147,* 1005-1022 doi.org/10.1037/xge0000449

Handel, S. (1989). *Listening: An introduction to the perception of auditory events.* Cambridge, Ma.: The MIT Press.

Healey, A. N., Sevdalis, N., & Vincent, C. A. (2006). Measuring intra-operative interference from distraction and interruption observed in the operating theatre. *Ergonomics, 49*, 589-604. doi.org/10.1080/00140130600568899

Hirsh, I. J. (1948a). Binaural summation and interaural inhibition as a function of the level of masking noise. *American Journal of Psychology*, *61*, 205-218, doi.org/10.2307/1416966

Hirsh, I. J. (1948b). The influence of interaural phase on interaural summation and inhibition. *Journal of the Acoustical Society of America, 20*, doi: 10.1121/1.1906407

Hodgetts, H.M., Farmer, E.W., Joose, M., Parmentier, F., Schaefer, D., Hoogeboom, P., van Gool, M. & Jones, D. (2005). The effects of party line communication on flight task performance. In: D. de Waard, K. Brookhuis, R. van Egmond, T. Boersema (Eds.), *Human factors in design, safety, and management.* Maastricht, NL: Shaker Publishing, pp. 327-338

Horswill, M. S., & McKenna, F. P. (1999). The effect of interference on dynamic risk-taking judgments. *British Journal of Psychology*, *90*, 189-199. doi.org/10.1348/000712699161341

Hughes, R. W. (2014). Auditory distraction. A duplex mechanism account. *PsyCH Journal, 3*, 30-41. doi.org/10.1002/pchj.44

Hughes, R. W., Hurlstone, M. J., Marsh, J. E., Vachon, F., & Jones, D. M. (2013). Cognitive control of auditory distraction: Impact of task difficulty, foreknowledge, and working memory capacity supports duplex-mechanism account. *Journal of Experimental Psychology: Human Perception & Performance*, *39*, 539-553 doi.org/10.1037/a0029064

Hughes, R. W., Vachon, F., & Jones, D. M. (2007). Disruption of short-term memory by changing and deviant sounds: Support for a duplex-mechanism account of auditory distraction. *Journal of Experimental Psychology: Learning, Memory & Cognition*, *33*, 1050-1061. doi.org/10.1037/0278-7393.33.6.1050

Jahncke, H., Hygge, S., Halin, N., Green, A. M., & Dimberg, K. (2011). Open-plan office noise: Cognitive performance and restoration. *Journal of Environmental Psychology, 31*, 373-382. doi.org/10.1016/j.jenvp.2011.07.002

James, W. (1890). *Principles of psychology*. New York: Holt.

Jones, D. M. (1995). The fate of the unattended stimulus: Irrelevant speech and cognition. *Applied Cognitive Psychology, 9,* 23-38. doi.org/10.1002/acp.2350090704

Jones, D. M. (1999). The cognitive psychology of auditory distraction: The 1997 Broadbent lecture. *British Journal of Psychology, 90*, 167-187. doi.org/10.1348/000712699161314

Jones, D. M., Alford, D., Bridges, A., Tremblay, S., & 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, & Cognition, 25,* 464–473. doi.org/10.1037/0278-7393.25.2.464

Jones, D. M., & Macken, W. J. (1993). Irrelevant tones produce an irrelevant speech effect: Implications for phonological coding in short-term memory. *Journal of Experimental Psychology: Learning, Memory & Cognition, 19,* 369-381. doi.org/10.1037/0278-7393.19.2.369

Jones, D. M., & Macken, W. J. (1995). Auditory babble and cognitive efficiency: Role of number of voices and their location. *Journal of Experimental Psychology: Applied, 1,* 216-226. doi.org/10.1037/1076-898X.1.3.216

Jones, D. M., Macken, W. J., & Murray, A. C. (1993). Privileged access by irrelevant speech to short-term memory: The role of changing state. *Memory & Cognition, 21*, 318-328. doi.org/10.1080/14640749208401304

Jones, D, M., Miles, C., & Page, J. (1990). Disruption of proofreading by irrelevant speech: Effects of attention, arousal or memory? *Applied Cognitive Psychology, 4*, 89-108. doi.org/10.1080/14640749208401304

Jones, D. M., & Morris, N. (1992). Irrelevant speech and serial recall: Implications for theories of attention and working memory. *Scandinavian Journal of Psychology, 33,* 212-229. doi.org/10.1111/j.1467-9450.1992.tb00911.x

Kattner, F., & Ellermeier, W. (2020). Distraction at the cocktail party: Attenuation of the irrelevant speech effect after a training of auditory selective attention. *Journal of Experimental Psychology: Human Perception & Performance, 46,* 10-20. doi: 10.1037/xhp0000695

Lavie, N. (1995). Perceptual load as a necessary prerequisite for selective attention. *Journal of Experimental Psychology: Human Perception & Performance, 21,* 441-468. doi.org/10.1037/0096-1523.21.3.451

Lavie, N., & Tsal, Y. (1994). Perceptual load as a major determinant of the locus of selection in visual attention. *Perception & Psychophysics, 56*, 183-197. doi.org/10.3758/BF03213897

Lee, A. K. C., Larson, E., Maddox, R. K., & Shinn-Cunningham, B. G. (2014). Using neuroimaging to understand the cortical mechanisms of auditory selective attention. Hearing Research, 307, 111-120. doi.org/10.1016/j.heares.2013.06.010

Licklider, J. C. R. (1948). The influence of interaural phase relations upon the masking of speech by white noise, *Journal of the Acoustical Society of America*, *20*, 150-159. doi.org/10.1121/1.1906358

Macken, W. J., Tremblay, S., Houghton, R. J., Nicholls, A. P., & Jones, D. M. (2003). Does auditory streaming require attention? Evidence from attentional selectivity in short-term memory. *Journal of Experimental Psychology: Human Perception & Performance, 2*, 43-51. doi.org/10.1037/0096-1523.29.1.43

Marsh, J. E., Hughes, R. W., & Jones, D. M. (2008). Auditory distraction in semantic memory: A process-based approach. *Journal of Memory and Language 58*, 682-700 doi.org/10.1016/j.jml.2007.05.002

Molly, K., Griffiths, T. D., Chait, M., & Lavie, N. (2015). Inattentional deafness: Visual load leads to time-specific suppression of auditory evoked responses. *Journal of Neuroscience, 35*, 16046-16054 doi.org/10.1523/JNEUROSCI.2931-15.2015

Moray, N. (1959) Attention in dichotic listening: Affective cues and the influence of instructions. *Quarterly Journal of Experimental Psychology*, *11*, 1156–11560. doi.org/10.1080/17470215908416289

Murphy, G., Groeger, J. A., & Greene, C. M. (2016). Twenty years of load theory: Where are we now and where should we go next? *Psychonomic Bulletin & Review, 23,* 1316-1340. doi.org/10.3758/s13423-015-0982-5

Murphy, S., Fraenkel, N., & Dalton, P. (2013) Perceptual load does not modulate auditory distractor processing. *Cognition*, *129,* 345-355 doi.org/10.1016/j.cognition.2013.07.014

Näätänen, R. (1990). The role of attention in information processing as revealed by event-related potentials and other brain measures of cognitive function. *Behavioral & Brain Sciences, 13, 201-288.* doi.org/10.1017/S0140525X00078407

Parmentier, F. B. R., Elsley, J. V., Andres, P., & Barcelo, F. (2011) Why are auditory novels distracting? Contrasting the roles of novelty, violation of expectation and stimulus change *Cognition*, *119*, 374-380. doi.org/10.1016/j.cognition.2011.02.001

Rock, I., & Guttman, D. (1981). The effect of inattention on form perception. *Journal of Experimental Psychology: Human Perception & Performance, 7,* 275-285. doi.org/10.1037/0096-1523.7.2.275

Salamé, P. & Baddeley, A. D. (1982). Disruption of short-term memory by unattended speech Implications for the structure of working memory. *Journal of Verbal Learning and Verbal Behavior*, *21*, 150-164 doi.org/10.1016/S0022-5371(82)90521-7

Scott, S. K., Rosen, S., Beaman, C. P., Davis, J. P. & Wise, R. J. S. (2009) The neural processing of masked speech: evidence for different mechanisms in the left and right temporal lobes. *Journal of the Acoustical Society of America*, *125*, 1737-1743 doi.org/10.1121/1.3050255

Shallice, T., & Cooper, R. (2011). *The organisation of mind*. Oxford: Oxford University Press.

Shield, B. M., & Dockrell, J. E. (2003). The effects of noise on children at school: A review. *Building Acoustics, 10*, 97-116. doi.org/10.1260/135101003768965960

Shinn-Cunningham, B. G. (2008). Object-based auditory and visual attention. *Trends in Cognitive Sciences*, *12*, 182-186. doi.org/10.1016/j.tics.2008.02.003

Simons, D. J., & Chabris, C. F. (1999). Gorillas in our midst: sustained inattentional blindness for dynamic events. *Perception*, *28*, 1059-1074. doi.org/10.1068/p281059

Sörqvist, P., Marsh, J. E., & Jahncke, H. (2010). Hemispheric asymmetries in auditory distraction. *Brain & Cognition*, *74*, 79-87. doi.org/10.1016/j.bandc.2010.06.007

Strayer, D. L., & Johnston, W. A. (2001). Driven to distraction: Dual-task studies of simulated driving and conversing on a cellular telephone. *Psychological Science*, *12*, 462–466. doi.org/10.1111/1467-9280.00386

Treisman, A. M. (1960). Contextual cues in selective listening. *Quarterly Journal of Experimental Psychology*, *12*, 242–248 doi.org/10.1080/17470216008416732

Treisman, A. M. (1964). Verbal cues, language, and meaning in selective attention. *The American Journal of Psychology, 77,* 206-219. doi: 10.2307/1420127

Tremblay, S., Parmentier, F. B. R., Hodgetts, H. M., Hughes, R. W., & Jones, D. M. (2012). Disruption of verbal-spatial serial memory by extraneous air-traffic speech. *Journal of Applied Research in Memory & Cognition, 1,* 73-79. doi.org/10.1016/j.jarmac.2012.04.004

Underwood, G. (1974). Moray vs. The rest: The effects of extended shadowing practice. *Quarterly Journal of Experimental Psychology*, *26*, 368–372 doi.org/10.1080/14640747408400426

Von Wright, J. M., Anderson, K., & Stenman, U. (1975). Generalization of conditioned GSRs in dichotic listening. In P. M. A. Rabbitt & S. Dornic (Eds.), *Attention and performance* (Vol. V, pp. 194–204). London: Academic Press.

Wood, N., & Cowan, N. (1995a). The cocktail party phenomenon revisited: Attention and memory in the classic selective listening procedure of Cherry (1953). *Journal of Experimental Psychology: General*, *124*, 243-262 doi.org/10.1037/0096-3445.124.3.243

Wood, N., & Cowan, N. (1995b). The cocktail party phenomenon revisited: How frequent are attention shifts to one's name in an irrelevant auditory channel? *Journal of Experimental Psychology: Learning, Memory, & Cognition, 21,* 255-260. doi:10.1037//0278-7393.21.1.255