Gorillas we have missed: Sustained inattentional deafness for dynamic events

Polly Dalton & Nick Fraenkel

Department of Psychology, Royal Holloway, University of London, UK

Abstract: 124 words

Main text (including abstract, but excluding title, author affiliations, key words, highlights, acknowledgments, bibliography and figures): 3893 words

Address correspondence to:

Polly Dalton, Department of Psychology,

Royal Holloway, University of London,

Egham, Surrey, TW20 0EX, UK.

Email: polly.dalton@rhul.ac.uk

Tel: (44) 1784 443516

Fax: (44) 1784 434347

Abstract

It is now well-known that the absence of attention can leave us 'blind' to visual stimuli that are very obvious under normal viewing conditions (e.g. a person dressed as a gorilla; Simons & Chabris, 1999). However, the question of whether hearing can ever be susceptible to such effects remains open. Here, we present evidence that the absence of attention can leave people 'deaf' to the presence of an 'auditory gorilla' which is audible for 19 seconds and clearly noticeable under full attention. These findings provide the first ever demonstration of sustained inattentional deafness. The effect is all the more surprising because it occurs within a lifelike, three-dimensional auditory scene in which the unnoticed stimulus moves through the middle of several other dynamic auditory stimuli.

Key words

Inattentional deafness; auditory attention; inattention; auditory awareness

Highlights

- The absence of attention can leave us 'blind', even to salient visual stimuli
- We tested whether hearing is susceptible to similar inattention effects
- We presented a lifelike, three-dimensional auditory scene containing several stimuli
- We tested detection of a salient auditory 'gorilla' stimulus, presented for 19 seconds
- Inattention caused many people to miss this salient auditory stimulus

Selective attention allows us to focus on some stimuli at the expense of others and is therefore crucial in allowing us to behave effectively in a complicated world. Some of the most striking demonstrations of the potential power of selective attention stem from studies of inattentional blindness (e.g. Mack & Rock, 1998) which have shown that the absence of attention can leave people ‘blind’ to visual stimuli that are obvious under normal viewing conditions (e.g. a person dressed as a gorilla; Simons & Chabris, 1999). The strength of these effects is likely to relate to the fact that the inattentional blindness paradigm examines people’s awareness of stimuli that are completely unexpected and therefore genuinely unattended. By contrast, in much other attention research, the unattended stimuli are presented throughout the experiment, raising the possibility that they might in fact receive some level of deliberate attentional allocation. In this respect, it might be argued that studies of inattentional blindness apply more directly to real world situations, in which stimuli arrive unpredictably and without advance warning. Perhaps for this reason, inattentional blindness has received large amounts of recent research interest. However the phenomenon has rarely been investigated in audition. Nevertheless, given the prevalent view of hearing as an ‘early-warning system’, tuned to detect unexpected stimuli (e.g. Scharf, 1998), it is important to ask whether inattention can have similarly pronounced effects on auditory awareness as it does on visual awareness.

Demonstrations of sustained inattentional blindness are typically elicited within a visual scene containing three separable elements: the task-relevant stimuli (e.g. the basketball players wearing white shirts in Simons and Chabris’s (1999) demonstration); the task-irrelevant stimuli (e.g. the players wearing black shirts); and the unexpected ‘critical’ stimulus (e.g. the person in the gorilla suit). The relevant and irrelevant stimuli are presented throughout the scene and differ from one another on one salient dimension (colour in the example above). The critical stimulus arrives without warning later in the scene and is similar to the irrelevant stimuli on the dimension that differentiates them from the relevant stimuli (i.e. being black in colour, in the above example) while nevertheless being clearly differentiable on a range of other dimensions (e.g. spatial location, speed, trajectory, shape etc.). The central finding is that the similarity between the unexpected critical stimulus and the irrelevant stimuli on the dimension upon which relevant and irrelevant stimuli are defined (i.e. colour in the current example) can prevent detection of the critical stimulus, despite its salience on a number of other dimensions (e.g. Most, Scholl, Clifford, & Simons, 2005; Simons & Chabris, 1999). Indeed, one experiment using audiovisual stimuli has shown that the strength of the visual selective attention effect achieved using this set-up can cause the critical stimulus to be missed even when it has a high level of auditory salience (Wayand, Levin & Varakin, 2005). However, it remains an open question as to whether auditory inattention can lead to similarly pronounced failures of awareness.

It is of course well-known that auditory attention can focus successfully on one stream at the expense of another. For example, in the influential dichotic listening paradigm, participants are presented with two spoken messages, one to each ear, and asked to attend to one while ignoring the other. The central focus of these studies has been the extent to which the unattended message is processed in the absence of focal attention. Although the debate continues over the exact extent of processing that the unattended message receives (e.g. Rivenez et al., 2008), the consensus is that genuinely unattended information is processed only to a relatively basic level (e.g. Broadbent, 1958;

Cherry, 1953; Holender, 1986; Lachter, Forster & Ruthruff, 2004). This argument was supported by a preliminary study of inattentional deafness using a dichotic listening set-up, which demonstrated that participants could fail to notice very brief, neutral stimuli presented to the unattended ear (Mack & Rock, 1998). However, although these studies demonstrate that information arriving at one ear can be prioritised at the expense of information arriving at the other ear, to our knowledge no one has yet tested whether a clearly noticeable and sustained auditory element can be introduced to the scene without detection, as has been demonstrated within the visual domain. Nevertheless, this is an important question, because the appearance without warning of new and unexpected scene elements constitutes a different type of potentially distracting input than the continuous presence of ongoing task-irrelevant information. Indeed, in real world situations, processing of new and unexpected stimuli (e.g. fire alarms, unexpected movement) is likely to be more important than processing of continually-present yet task-irrelevant scene elements (e.g. background conversations).

Presumably, auditory attention research has focused on tasks involving only two elements at least in part because of the relative ease of delivering two clearly separable messages by presenting one to each ear. Here, by contrast, we use binaural recording and presentation to create a naturalistic, three-dimensional auditory scene consisting of multiple, dynamic stimuli. Our use of a binaural scene is important for creating an analogue of the inattentional blindness paradigm, as it allows us to present three separable scene elements. However, the technique also allows us to test the possibility that the dichotic listening research might have over-estimated the effects of auditory selective attention by presenting the attended and unattended streams at fixed and easily-separable locations. Indeed, the presentation of one message to each ear delivers the most extreme level of spatial separation possible within audition. It is thus likely to be easier to focus attention selectively in a dichotic listening experiment than in a dynamic, three-dimensional auditory scene. Our set-up allows for the important test of whether the strong selective attention effects revealed within the dichotic listening paradigm can extend to a context that affords much less extreme spatial separation between messages.

We created a binaural scene containing one conversation between two women presented at the same time as a different conversation between two men. Halfway through this scene, an additional male character walked through the room, continually repeating the phrase “I’m a gorilla” (after Simons & Chabris, 1999) for 19 seconds. Half of the participants were asked to listen to the women’s conversation. In this case, the gorilla stimulus was clearly separated in space from the attended conversation (see Figure 1) and, being a man’s voice, fell into an unattended category of stimulus – a situation which has been shown to lead to inattentional blindness in the visual domain. Our central question was whether participants in this condition would experience ‘inattentional deafness’, such that they might fail to notice the gorilla. The remaining participants attended to the men’s conversation. In this case, the gorilla stimulus fell into the attended category and passed closer to the attended conversation, and we therefore predicted that most participants would notice it.

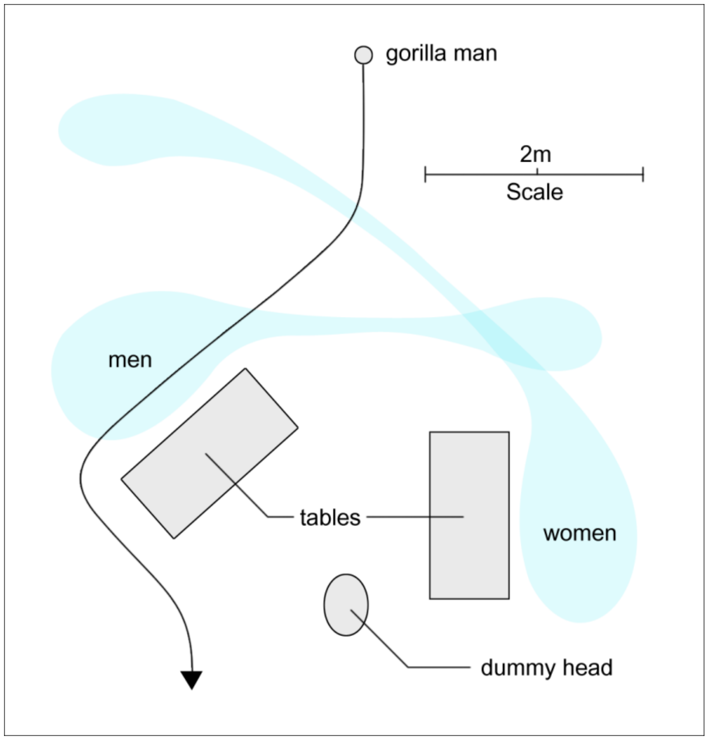

Figure 1. Schematic depicting the trajectory of the auditory ‘gorilla’ in relation to the other elements of the scene. Note that half the participants heard a reverse stereo version of this scene, in which the left and right channels were swapped (so that the men were situated mainly to the right of the head and the gorilla also passed through that side).

1. Experiment 1

1.1 Methods

1.1.1 Participants

45 people aged between 16 and 47 (mean age 20) participated, either voluntarily, as part of a research demonstration (in the case of visiting school groups), or in return for £2 (when recruited from the Royal Holloway campus). Informed consent was obtained and all aspects of the protocol were approved by the Departmental Ethics Committee.

1.1.2 Materials

We presented listeners with an auditory scene, recorded binaurally using a dummy head in order to convey a realistic sense of three-dimensional auditory space. This provided our experimental stimuli with an unusually high degree of realism (in order to listen to the audio from this experiment, please visit: http://www.pc.rhul.ac.uk/sites/attentionlab/demo/gorilla/). [NB We would ask that this link is only made available to relevant editors and reviewers in advance of publication]. The scene lasted for 69 seconds and contained four characters preparing for a party: two women wrapping up a present and two men preparing food and drink. All characters moved around the room during the scene, though the men spent most of their time to the left of the head while the women were mostly to the right. After 33 seconds, an additional male character entered from the back of the room and walked through the scene, passing by the left of the head, continually repeating the phrase "I'm a gorilla". The gorilla was audible for 19s and his trajectory is illustrated in Figure 1. The gorilla was recorded in a separate take from the rest of the scene, and mixed in subsequently, in order to allow independent manipulations of this stimulus (cf. Experiment 2). In order to balance for any potential stereo orientation effects, half of the participants listened to the scene with the right and left channels reversed, creating a 'mirror image' version of the scene. All experimental instructions were spoken in a female voice and presented with quiet background music, in order to distinguish them from the binaural scenes.

1.1.3 Apparatus

The binaural recording was produced using a dummy head – designed to replicate as closely as possible the acoustic properties of a real human head – modelled around a plastic resin skull, with moulded rubber latex flesh and a dense foam 'brain'. The ears and other facial features were moulded from human casts. Inside each ear was a DPA

4060 omnidirectional hi-sensitivity condensor microphone. The experimental instructions were recorded with a standard dynamic microphone (Behringer XM8500).

The audio was recorded via an Edirol UA-25 USB soundcard at 48kHz / 24-bit to a laptop running Cockos REAPER digital audio workstation software, and all subsequent editing was carried out in REAPER at 48kHz / 24-bit, and mixed down to 44kHz / 16-bit for use in the experiment.

The experiment was programmed in PST E-Prime 2.0. Because testing occurred in several different locations, a variety of PCs, headphones and microphones (for recording participants' responses) was used.

1.1.4 Design and procedure

Participants proceeded through the experiment at their own pace, pressing SPACE to initiate each stage. First, they were played an example binaural recording, in order to familiarise them with the extremely realistic sense of space that such recordings convey.

In the experimental phase, all participants were asked to listen carefully to the binaural 'gorilla' scene in order to answer some subsequent questions. Half of the participants were asked to listen to the women's conversation and the other half were asked to attend to the men's conversation. It is important to note that the two groups listened to exactly the same scene, meaning that any differences in detection rates between the groups can be attributed directly to differences in participants' attentional focus.

Following presentation of the scene, participants were asked two questions: (1) Did you hear anything unusual that didn't fit in with the scene? (2) Did you hear anyone other than the four people preparing for the party? They responded to these questions by pressing ‘Y’ for ‘yes’ or ‘N’ for ‘no’. Following a ‘yes’ response to either question, participants were asked to give more details about what they had heard. Participants were then able to record their own spoken responses via a microphone connected to the computer (pressing SPACE to start and stop recording).

Next, in order to confirm that all participants could hear the gorilla under conditions of full attention, we asked them to listen to the scene again, this time attending to the men’s voices and listening for ‘anything unusual’. At the end of the scene, they were asked the same questions and gave responses in the same way as described above.

The screen remained blank throughout, other than for the ‘welcome’ and ‘goodbye’ screens, and for on-screen reminders where a key-press response was required. The experiment took less than 10 minutes to complete and finished with a full, spoken debrief, pre-recorded and presented at the end of the experiment.

1.2 Results and discussion

The data from four participants were excluded due to technical difficulties with the experimental procedure. The data from another participant were removed because of a failure to detect the gorilla stimulus on the control trial. (Note that the fact that only one out of 45 participants failed to notice the gorilla stimulus on the control trial indicates that the stimulus was clearly audible under conditions of full attention).

The remaining 40 participants were distributed equally between both attention conditions (Attend Women vs. Attend Men) and stereo orientations (Normal vs. Reverse). There were no effects of stereo orientation on responding, ($\chi 2$ (1, N = 40) = 1.67, $p > .10$) so we combined the groups for subsequent analyses.

Table 1 indicates the stage at which participants in each condition first reported detecting the unexpected stimulus. In the 'Attend Men' condition, 90% of participants spontaneously mentioned the gorilla in response to the first question. This demonstrates that the gorilla was clearly noticeable to participants who were paying attention to male voices. By contrast, significantly fewer participants (30%) in the 'Attend Women' condition noticed the gorilla on first hearing the scene ($\chi 2$ (1, N = 40) = 15, $p < 0.001$). Thus 70% of people in this condition remained unable to report the presence of a man repeating "I'm a gorilla", even though he remained audible for 19 seconds and walked right through the middle of a realistic auditory scene.

	Response		
	Reported gorilla on inattention trial, Q1 ("Anything unusual?")	Noticed voice but not gorilla on inattention trial, Q2 ("Anyone other than people preparing for party?")	Only noticed gorilla on full attention control trial
Attend women	6	2	12
Attend men	18	0	2

Table 1. Numbers of participants in each condition noticing the unexpected stimulus at each stage of Experiment 1.

Our results show that the absence of attention can leave people 'deaf' to a sustained and dynamic auditory stimulus that is clearly noticeable under normal listening conditions, providing the first ever demonstration of sustained inattentional deafness. These findings illustrate the potential power of auditory selective attention in a context that is more realistic and challenging than the dichotic listening tasks which have typically been used in this research area.

However, whereas the dichotic listening research demonstrated that very little of an ongoing unattended auditory stream is processed, the focus of the present study was somewhat different. Here, we set out to test people's detection of an additional, sustained scene element that appeared without warning alongside the ongoing attended and unattended streams. The inattentional deafness that participants showed with respect to this unexpected stimulus – with the vast majority completely unable to report its presence – appears to constitute a qualitatively different phenomenon from that associated with the lack of processing of the unattended stream in a dichotic listening set-up – where participants remain aware of the presence and basic physical characteristics of the unattended stream, even though they are typically unable to report its semantic content (e.g. Cherry, 1953).

It is likely that the auditory scene was segmented differently in the 'Attend Men' and 'Attend Women' conditions (see, e.g., Bregman, 1994). Participants in the 'Attend Men' condition were able to perceive the attended men's conversation and the male gorilla as separate scene elements. It thus seems likely that the men's conversation constituted one stream and the gorilla, when it appeared, constituted a second stream, to which 90% of participants allocated some attention, presumably because it possessed task-relevant features (e.g. Shinn-Cunningham, 2008). It is less clear how auditory scene analysis might have proceeded in the 'Attend Women' condition, in which both the men's conversation and the gorilla were unattended. The question of whether auditory streams can be formed in the absence of attention remains open, with some research suggesting that this is possible (Sussman et al., 2007) and other studies finding the opposite (Cusack et al., 2004). It is thus possible that, in the absence of attention, the gorilla was not segmented from the men's conversation, despite being distinct on a number of dimensions (e.g. trajectory, timing of utterances and prosody). In this case, the inattentional deafness seen within the 'Attend Women' condition might reflect a failure to notice the addition of an unexpected speaker to an unattended stream, rather than a failure to notice an entirely new stream – a finding which might be comparable to earlier demonstrations indicating limited processing of unattended dichotic listening streams. We note, however, that the addition of a new speaker would cause large changes in what Cherry (1953, pg. 978) called the "statistical properties" of the message, which would typically be identified even within the unattended stream of a dichotic listening task. From this perspective, the sustained inattentional deafness seen in Experiment 1 would be surprising even if the gorilla did appear within an unattended stream (particularly given that our stimuli were presented within a dynamic auditory scene in which attended and unattended elements moved around each other – a set-up which might if anything have been expected to produce less pronounced attentional effects than the dichotic listening technique, in which attended and unattended stimuli were presented at fixed and easily separable locations).

Nevertheless, in order to investigate this issue further, we ran a second experiment in which the gorilla stimulus was presented in 'mirror image' with respect to the other stimuli, such that it now appeared on the opposite side of the scene. This was designed to increase the segregation between the gorilla and the men's conversation, by separating them clearly in space. The resulting scene also provides an interesting comparison with the dichotic listening research, because any finding of inattentional deafness in the 'Attend Women' condition will now indicate a failure to detect a salient auditory stimulus passing right through the attended side of the scene, akin to an unexpected speaker appearing in the *attended* ear of a dichotic listening task.

2. Experiment 2

2.1 Methods

2.1.1 Participants

50 people aged between 16 and 60 (mean age 27) participated. Recruitment and payment conditions were as described for Experiment 1.

2.1.2 Apparatus, design, materials and procedure

All aspects of the method were identical to those of Experiment 1, with the sole difference that the right and left channels of the recording of the 'gorilla' stimulus were reversed, producing a 'mirror image' version of that stimulus which was then mixed into the original auditory scene. This had the effect of switching the gorilla stimulus from predominantly occupying the same side of space as the men's conversation (as shown in Figure 1) to predominantly occupying the same side of space as the women.

2.2 Results and discussion

The data from four participants were excluded due to technical difficulties with the experimental procedure. The data from a further six participants were removed because of a failure to detect the gorilla stimulus on the full attention control trial.

The remaining 40 participants were distributed equally between both attention conditions (Attend Women vs. Attend Men) and stereo orientations (Normal vs. Reverse). There were no effects of stereo orientation on responding, ($\chi 2$ (1, N = 40) = 1.62, $p > .10$) so we combined the groups for subsequent analyses.

Table 2 indicates the stage at which participants in each condition first reported detecting the unattended stimulus. In the 'Attend Men' condition, 65% of participants spontaneously mentioned the gorilla in response to the first question. Between-experiment comparisons indicated that participants in this condition demonstrated marginally more inattentional deafness than those in the comparable condition of Experiment 1 ($\chi2$ (1, N = 40) = 3.58, p = .059), presumably because the gorilla no longer appeared at an attended location.

Fewer participants (45%) in the 'Attend Women' condition noticed the gorilla on first hearing the scene, although the difference between the two conditions was not significant ($\chi2$ (1, N = 40) = 1.62, p > .10). This lack of difference is not surprising, given that the gorilla now appeared at a task-relevant location in the 'Attend Women' condition and with a task-relevant voice in the 'Attend Men' condition. By contrast, in Experiment 1 the gorilla's spatial location and voice category were both task-relevant in the 'Attend Men' condition and task-irrelevant in the 'Attend Women' condition. Indeed, the lack of a difference between the conditions in Experiment 2 is interesting in suggesting that both spatial location and voice category influenced attentional allocation within the scene, rather than one of these cues being dominant.

	Response		
	Reported gorilla on inattention trial, Q1 ("Anything unusual?")	Noticed voice but not gorilla on inattention trial, Q2 ("Anyone other than people preparing for party?")	Only noticed gorilla on full attention control trial
Attend women	9	0	11
Attend men	13	2	5

Table 2. Numbers of participants in each condition noticing the unexpected stimulus at each stage of Experiment 2.

However, the most important finding with respect to the current research question was that 55% of people in the 'Attend Women' condition failed to notice the gorilla. Indeed, between-experiment comparisons indicated that there was no significant difference in the extent of inattentional deafness seen in the 'Attend Women' conditions in Experiments 1 and 2 ($\chi 2$ (1, N = 40) = 0.96, p > .10). The majority of people failed to notice an unexpected stimulus that moved through the centre of the conversation to which they were attending – the equivalent of a new voice appearing within the *attended* ear in a dichotic listening task.

3. General discussion

The present experiments show that the absence of attention can leave people 'deaf' to a sustained and dynamic auditory stimulus that is clearly noticeable under normal listening conditions, providing the first ever demonstration of sustained inattentional deafness. These findings illustrate the potential power of auditory selective attention in a realistic and dynamic context where segregation of relevant and irrelevant stimuli is far more challenging than in the dichotic listening tasks typically used in this research area.

Our findings add to a growing body of literature demonstrating the importance of attention in determining auditory awareness. For example, studies of the auditory attentional blink have shown that people can fail to detect the second of two auditory targets if it appears before processing of the first target has finished (e.g. Duncan, Martens & Ward, 1997; Mondor, 1998). However the failures of auditory awareness demonstrated within that paradigm typically last for around half a second and depend on a very precise experimental set-up (e.g. Tremblay, Vachon & Jones, 2005). By contrast, in the current study, a distinctive auditory stimulus remains undetected for 19 seconds within a lifelike, dynamic auditory scene. Research into ‘change deafness’ has also demonstrated failures of auditory awareness, showing that a change introduced between two successive versions of an auditory scene can go undetected in the absence of directed attention to the changing feature (e.g. Eramudugolla et al., 2005; Vitevich, 2003). However, these failures of awareness relate to difficulties in comparing successive representations of auditory scenes. By contrast, the current findings reflect an ongoing lack of awareness for a distinctive yet unexpected element within a single scene.

In summary, this study establishes the new phenomenon of sustained inattentional deafness. The findings demonstrate the power of auditory selective attention to prioritise relevant over irrelevant information, even within a dynamic, three-dimensional scene. The fact that inattention can cause us to miss surprising and potentially important auditory stimuli also has significant real-world implications in areas such as road safety and interface design.

4. Acknowledgements

This research was funded by an Economic and Social Research Council (UK) small research grant (RES-000-22-3365) to P.D. We thank Sian Beilock and three anonymous reviewers for their comments on an earlier version of this manuscript. We are also grateful to the Adult and Child Cognition group at Royal Holloway for helpful discussions and to Oliver Samuel for providing us with the dummy head.

Correspondence concerning this article should be addressed to Polly Dalton, Department of Psychology, Royal Holloway University of London, Egham, Surrey, TW20 0EX, UK. Electronic mail may be sent to polly.dalton@rhul.ac.uk.

5. References

Bregman, A. S. (1994). *Auditory scene analysis: the perceptual organization of sound.* MIT Press.

Broadbent, D. E. (1958). *Perception and Communication*. New York: Oxford University Press.

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.

Cusack, R., Deeks, J., Aikman, G., & Carlyon, R.P. (2004). Effects of location, frequency region, and time course of selective attention on auditory scene analysis. *Journal of Experimental Psychology: Human Perception and Performance*, *30*, 643-656

Duncan, J., Martens, S., & Ward, R. (1997). Restricted attentional capacity within but not between sensory modalities. *Nature, 387*, 808–810.

Eramudugolla, R., Irvine, D. R. F., McAnally, K. I., Martin, R. L., & Mattingley, J. B. (2005). Directed attention eliminates 'change deafness' in complex auditory scenes. *Current Biology, 15*, 1108–1113.

Holender, D. (1986). Semantic activation without conscious identification in dichotic listening, parafoveal vision, and visual masking: A survey and appraisal. *Behavioral and Brain Sciences, 9*, 1-23.

Lachter, J., Forster, K. I., & Ruthruff, E. (2004). Forty-Five Years After Broadbent (1958): Still No Identification Without Attention. *Psychological Review, 111*, 880-913.

Mack, A., & Rock, I. (1998). *Inattentional Blindness.* MIT Press.

Mondor, T. (1998). A transient processing deficit following selection of an auditory target. *Psychonomic Bulletin & Review, 5*, 305-311.

Most, S. B., Scholl, B. J., Clifford, E. R., & Simons, D. J. (2005). What you see is what you set: sustained inattentional blindness and the capture of awareness. *Psychological Review, 112*, 217-242.

Rivenez, M., Guillaume, A., Bourgeon, L., and Darwin, C. J. (2008). Effect of voice characteristics on the attended and unattended processing of two concurrent messages. *Journal of Cognitive Psychology, 20*, 967-993.

Scharf, B. (1998). Auditory attention: The psychoacoustical approach. In H. Pashler (Ed.), *Attention* (pp. 75-117). Hove, UK: Psychology Press.

Shinn-Cunningham, B.G. (2008). Object-based auditory and visual attention. *Trends in Cognitive Sciences, 12*, 182-186.

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

Sussman, E.S., Horvath, J., Winkler, I., & Orr, M. (2007). The role of attention in the formation of auditory streams. Perception and Psychophysics, 69, 136-152.

Tremblay, S., Vachon, F., & Jones, D. M. (2005). Attentional and perceptual sources of the auditory attentional blink. *Perception & Psychophysics, 67*, 195–208.

Vitevitch, M. S. (2003). Change deafness: The inability to detect changes between two voices. *Journal of Experimental Psychology: Human Perception & Performance, 29,* 333–342.

Wayand, J. F., Levin, D. T., & Varakin, D. A. (2005). Inattentional blindness for a noxious multimodal stimulus. *The American Journal of Psychology, 118*, 339-352.