Object-based auditory and visual attention

Barbara G. Shinn-Cunningham

Hearing Research Center, Departments of Cognitive and Neural Systems and Biomedical Engineering, Boston University, Boston, MA 02215, USA
Speech and Hearing Bioscience and Technology Program, Harvard-MIT Division of Health Sciences and Technology, Cambridge, MA 02139, USA

Theories of visual attention argue that attention operates on perceptual objects, and thus that interactions between object formation and selective attention determine how competing sources interfere with perception. In auditory perception, theories of attention are less mature and no comprehensive framework exists to explain how attention influences perceptual abilities. However, the same principles that govern visual perception can explain many seemingly disparate auditory phenomena. In particular, many recent studies of 'informational masking' can be explained by failures of either auditory object formation or auditory object selection. This similarity suggests that the same neural mechanisms control attention and influence perception across different sensory modalities.

Introduction

At a cocktail party, the sounds of clinking glasses and exuberant voices add acoustically before entering your ears. To appreciate your companion's anecdote, you must filter out extraneous sources (see Glossary) and focus attention on her voice. At the same time, the sounds that you tune out are crucial for maintaining awareness of your environment. Indeed, a source of interference (the pompous man on your right) might become the very source you want to understand in the next moment (e.g. when you realize he is relaying a juicy story about your boss). To maneuver successfully in everyday settings, you need to be able to both focus and shift attention as the need arises.

Theories of visual attention explain many striking perceptual phenomena that arise when viewing complex scenes, from change blindness (failure to notice a change in a visual scene because attention is directed to another part of the image) to performance on visual search tasks [1,2]. Although there is much current interest in how central limitations interfere with auditory perception, there is no comprehensive framework to explain our ability to understand sound sources in complex acoustic scenes. Here, I argue that many auditory phenomena, including how we manage to converse at a cocktail party, can be understood by properly extending theories of visual attention. This commonality supports the idea that the same neural processes control visual and auditory attention [3].

Auditory objects

Theories of visual attention argue that observers focus attention on an object in a complex scene [2]. Unfortunately, just as in vision [4], it is difficult to define what constitutes an object in audition. This difficulty arises, in part, because there are few absolute rules governing auditory object formation. Audible sound in a mixture is not always allocated between the objects perceived in a scene, and can contribute either to multiple objects [5,6] or to no object [7]. The state of the listener, from expectations about a scene's content to the level of analysis a listener undertakes (listening to a symphony orchestra versus the English horn solo), influences the perceived content of an object [8,9]. Particularly for ambiguously structured stimuli, the perceptual organization of a scene evolves over time and/or is bistable [10,11].

Despite the lack of a precise definition, we have an intuitive understanding of what an auditory object is. At the cocktail party, we perceive the woman speaking on the left, the chiming doorbell, a shattering plate. Each of these auditory objects is an estimate of sound emanating from a discrete sound source: an 'auditory object' is a perceptual entity that, correctly or not, is perceived as coming from one physical source.

Object formation

In a visual scene, objects form locally based on contiguous geometric structure, such as edges, boundaries and contours [4]. Discrete local patches can be perceptually linked, based on similarity of texture, color and other features, to form whole objects [4].

In a similar way, auditory objects form across different analysis scales. For sound elements with contiguous

Glossary

Energetic masking: perceptual interference present in the sensory epithelium.
Informational masking: perceptual interference that cannot be explained by energetic masking.
Object: a perceptual estimate of the content of a discrete physical source.
Salience: the perceptual strength of an input based purely on stimulus attributes.
Similarity: a putative explanation for auditory informational masking when a target and competing sources have similar perceptual features.
Source: a discrete physical entity in the external world.
Streaming: grouping of short-term auditory objects across longer time scales.
Uncertainty: a putative explanation for auditory informational masking when properties of either the target or the masker change unpredictably from trial to trial.

Corresponding author: Shinn-Cunningham, B.G. (shinn@cns.bu.edu).

1364-6613/$ – see front matter © 2008 Elsevier Ltd. All rights reserved. doi:10.1016/j.tics.2008.02.003 Available online 7 April 2008

spectro-temporal structure, formation relies primarily on this local structure [12,13], including common onsets and offsets, harmonic structure and continuity of frequency over time. Owing to the physical constraints of how sound is produced, many ecological signals (particularly information-conveying communication signals, from birdsongs to speech) have a rich spectro-temporal structure that supports robust short-term object formation (e.g. formation of syllables). Short-term objects are streamed (linked together over time) through continuity and similarity of higher-order perceptual features, such as location, pitch, timbre and even learned meaning (word identity, grammatical structure, semantics) [13].

The relative influence of a particular cue or feature on object formation depends on the scale of the analysis. For instance, spatial auditory cues have a relatively weak influence over local time scales [13,14]. However, perceived location (as opposed to basic spatial cues, such as interaural time differences [15]) strongly influences how we link short-term auditory objects into a coherent stream [16].

Although the above description might seem to suggest that objects are constructed through a hierarchy of processing, first grouped based on local structure and second organized across longer spatial or temporal scales, the truth is more complex. Higher-order features and top-down attention can alter how objects form locally. Rather than a hierarchical processing structure, objects are formed through heterarchical interactions (between processes that mutually influence one another, rather than through a sequence of processing stages) across different scales. The ultimate perceptual organization of the scene, at all scales, depends on the preponderance of all evidence [7].

Object-based attention

Object formation directly influences how we perceive and process complex scenes. In all sensory modalities, the normal mode of analyzing a complex scene is to focus on one object while other objects are in the perceptual background [17,18]. In vision, this mode of perceiving is described as a biased competition between perceptual objects [2]. Biased competition takes place automatically and ubiquitously when there are multiple objects in a scene. Which object wins the competition depends both on the inherent salience of the objects and the influence of volitional, top-down attention, which biases the competition to favor objects with desired perceptual features [2,19].

Even when observers select what to attend based on low-level features, attention operates on objects [2,20]. For instance, when attention is spatially focused, the observers' sensitivity to other features that are part of the object at the attended location is also enhanced [17]. Thus, object formation is intricately linked with selective attention: the perceptual unit of attention is the object.

Most work on attention and objects is in the visual literature [2,21], but similar principles govern auditory perception [22–24]. Evidence suggests that attention acts on auditory objects, much as it enhances visual objects [25–27]. Moreover, listeners seem to attend actively to one, and only one, auditory object at a time [28,29], consistent with the biased-competition model of visual attention (Box 1).

Understanding perception of complex scenes

Because attention is object based, competing sources in a complex scene can cause many different forms of perceptual interference, some of which are considered below. An overview of the interactions affecting auditory perception is shown in Figure 1.

Energetic masking

The simplest form of perceptual interference occurs when a competing source renders portions of a target imperceptible. This kind of interference, known in auditory circles as energetic masking, occurs when the response on the sensory epithelium to the target is disrupted because the system is responding to a competing source. In the auditory domain, where the auditory nerve encodes sound in a time-frequency representation, energetic masking occurs when the masking signal overlaps in time and frequency with the target. In vision, an analogous form of interference occurs when a source near the observer obscures all, or part, of another source behind it. In such cases, the neural response to the target is distorted or imperceptible.

Current models can account for acoustic energetic masking effects; however, in some situations, performance is worse than predicted [30]. Many natural sounds, such as speech, are spectro-temporally sparse, so energetic masking often affects only portions of the target, limited in both time and frequency [31]. Moreover, we perceptually fill in inaudible portions based on glimpses we hear [32–34] (Box 2). As a result, energetic masking is often not the main factor limiting performance.

Box 1. Shifting attention

Evidence indicates that we listen to only one object at a time. Listeners have difficulty making judgments of the relative timing of events across (but not within) streams [12]. When listeners are asked to divide attention between two speech streams that are close together in space, they are able to report many of the words in the two streams, but intermingle words from the two messages [46]. By contrast, when the two streams are spatially distinct, listeners are less likely to confuse words across streams, but also recall fewer words overall [46]. These results hint that the more distinct competing streams are from one another, the more complete the suppression of the stream in the perceptual background.

How is it, then, that in everyday listening situations we seem to be able to understand multiple sources, especially in social settings where the flow of conversation is chaotic and unpredictable?

It is probable that we switch attention between objects in a complex setting, time-sharing attention between competing sources. Even if we do not perceive all of the content of one signal, we can fill in missing snippets (see also Box 2). In addition, we can use short-term sensory memory to help this filling-in process, mentally replaying the bit of the input signal that we did not focus on initially.

Switching attention takes in the order of 100–200 ms and sensory memory degrades with time. Thus, some of the information in a newly attended stream will be missed even after a listener switches attention. Moreover, auditory streams build-up over time [8,10], which might enhance the ability to focus on the stream in the perceptual foreground and understand its content. Thus, if listeners switch attention between streams, performance is likely to be degraded owing to the direct cost of switching attention and because switching attention resets streaming, negating the benefit of object build-up.

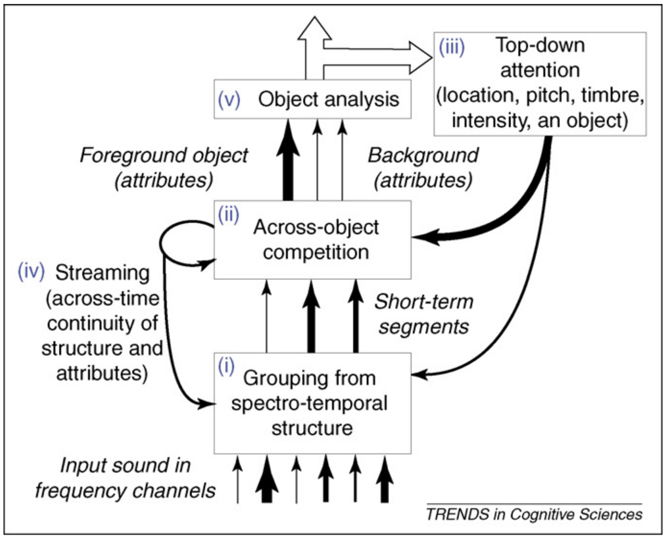
(iii) Top-down attention (location, pitch, timbre, intensity, an object)
(v) Object analysis
Foreground object (attributes)
Background (attributes)
(ii) Across-object competition
(iv) Streaming (across-time continuity of structure and attributes)
Short-term segments
(i) Grouping from spectro-temporal structure
Input sound in frequency channels
TRENDS in Cognitive Sciences

Figure 1. Conceptual model relating auditory object formation and its interactions with bottom-up salience and top-down attention, where arrow width denotes the strength of a signal or a connection. (**i**) Short-term segments initially form based on local spectro-temporal grouping cues [12,13]. (**ii**) Competition first arises between short-term segments. Some segments might be inherently more salient than others (e.g. because of their intensity or distinctiveness) [41,44], which biases the intersegment competition. (**iii**) Top-down attention and (**iv**) streaming (across-time linkage based on bottom-up object continuity) help modulate the competition, biasing it to favor objects with desirable features and to maintain attention on the object already in the foreground [23,45]. (**v**) As a result, one object is emphasized at the expense of others in the scene [46].

Informational masking

Perhaps because there is no widely accepted theory to explain auditory interference beyond energetic masking, the catchall phrase ‘informational masking’ is used to encompass all masking that is not energetic [30]. Although there is a large and growing interest in informational masking, mechanistic explanations are lacking (see also Box 3). Here, I argue that results of many studies of informational masking can be explained by failures of object-based attention.

Box 3. Understanding stimulus similarity and uncertainty

Many recent psychoacoustic studies link informational masking with stimulus similarity (i.e. similarity between target and maskers) and with stimulus uncertainty (e.g. randomness in the masker and/or target) [35,42]. Although similarity and uncertainty affect informational masking, here I argue that they do so by affecting object formation and object selection.

Similarity between target and masker can cause either or both of the processes of object formation and object selection to fail. Similarity can cause the target and masker to be perceived as part of the same, larger perceptual object, which will result in poorer sensitivity to the content of the target [36] (Figure 2a). Even if target and masker are perceptually segregated into distinct objects, similarity of these objects can interfere with the selection of the correct object in a scene.

Uncertainty also can interfere with object selection, either because the listener is unsure of how to direct top-down attention to select the target object [49] or because the salience of new events (e.g. randomly varying maskers) draws exogenous attention too strongly to be overcome by top-down attention [41].

Although stimulus similarity and uncertainty influence perception in a complex scene, the processes underlying these effects can be attributed to object-based auditory attention. Framed in this way, results from many different studies of informational masking can be understood and explained.

Failures of object formation

Failures in object formation can come about when local structure is insufficient to separate one source from others in a scene [35], which can degrade perception [36]. This can occur for a variety of reasons, including: 1) energetic masking might make all, or part, of the target imperceptible; 2) the mixture might contain competing sources that have similar spectro-temporal structure and that tend to group with the target; or 3) the target might not be structured enough to support object formation, for instance, if the mixture contains ambiguous or conflicting cues.

Figure 2 shows, by visual analogy, the kind of perceptual problems that can arise when local object formation breaks down. In the auditory domain, ‘double vowel’

Box 2. Coping with ambiguity: phonemic restoration

Picture yourself at a crowded bar with your friends. In this kind of setting, you can imagine a burst of laughter that momentarily masks your buddy’s unending tale of romantic misfortune. Fortunately (or perhaps unfortunately), speech signals are redundant and we can often understand a message even when we only hear glimpses of the speech signal [31]. Moreover, we perceptually fill in missing bits of speech based on the glimpses we hear (Figure I), so that we often do not even notice the interruption (an effect known as ‘phonemic restoration’) [47]. This ability depends on integrating all available evidence (including evidence for how to perceptually organize the scene) [34] to make sense of the message we want to understand. Thus, to make sense of noisy signals we hear in everyday settings, we depend on signal redundancy (from continuity of spectro-temporal energy in the sound to lexical, linguistic and semantic constraints) [47,48]. Although phonemic restoration is particularly strong for speech signals, even non-speech signals can be perceptually completed based on low-level spectro-temporal structure [32].

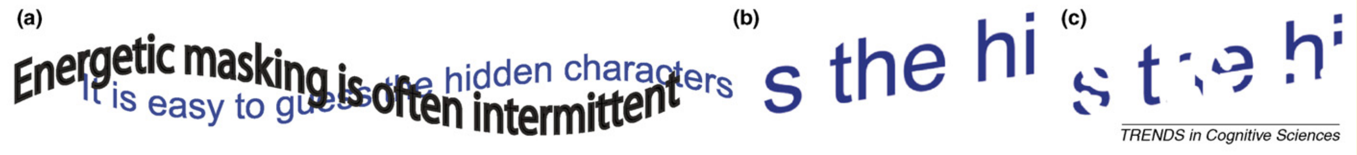
(a)
Energetic masking is often intermittent
It is easy to guess the hidden characters
(b)
s the hi
(c)
s t he h
TRENDS in Cognitive Sciences

Figure I. Visual analogy illustrating glimpsing and phonemic restoration. (**a**) Mixture of messages. Even though one message obstructs a portion of the other, the meaning of both messages is clear. Moreover, you undoubtedly perceive the full characters ‘the’ to be in the visual scene, even though the actual stimulus is ambiguous and could contain only portions of letters consistent with that interpretation. Your experience and knowledge enable you to perceptually fill in the hidden pieces based on what is most probable, given the sensory evidence you perceive as well as your knowledge of letters, words and meaning. (**b**) The center portion of the perceived text in the message beginning with ‘It is easy. . .’ comprises full characters, even though the other message beginning with ‘Energetic. . .’ occludes portions of the characters. (**c**) One possible center portion of the message beginning with ‘It is easy’ that would lead to the same physical stimulus reaching the observer’s retina as the messages shown in (a). Because of the occlusion by the message beginning with ‘Energetic. . .’, the hidden portions of the characters in the other message might not actually be physically present in the visual scene.

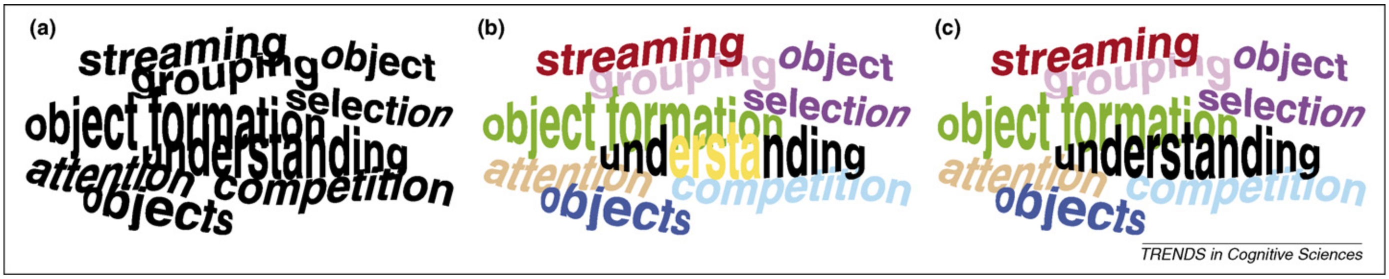

Figure 2. Visual analogies of failed object formation. (**a**) The general similarity of the features and elements of the image make it difficult to segregate words, so viewers are likely to perceive the mixture as a connected mass that fails to represent any of the individual words. When this occurs, it takes extra time and cognitive effort to understand the words. (**b**) When color is used to differentiate the letters, like-colored letters tend to group; however, if the letters making up the target word fail to group together and the target is not perceived as one unified object (direct attention to the middle of the image), analyzing the target word still requires extra effort. (**c**) Understanding is clear when the letters making up each word group together and each word forms automatically, resulting in an enhanced ability to selectively attend to each in turn.

experiments demonstrate failures of object formation: when two vowels are played with common onsets and offsets, listeners have difficulty identifying either vowel (local object formation fails because of ambiguous spectro-temporal structure); however, when harmonic structure differentiates the competing vowels, identification improves [37].

Failures in streaming occur when there are multiple sources that have similar higher-order features, such as when a listener hears a mixture of multiple male voices or the target is a set of tones amidst similar tones [38]. These failures can result in a target stream that is corrupted by sound elements from a masker or that is missing key elements (Figure 3), which can interfere with perception of the target.

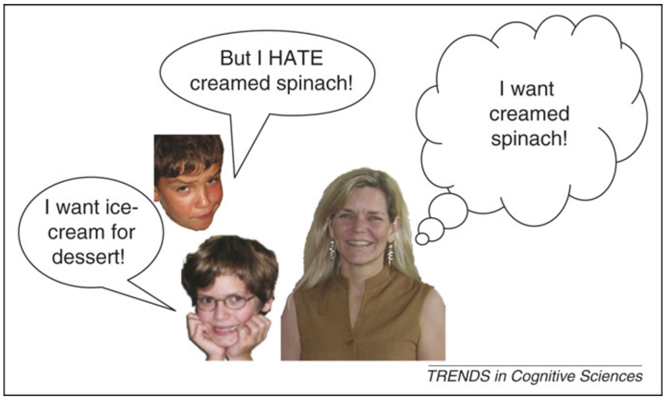
But I HATE creamed spinach!
I want creamed spinach!
I want ice-cream for dessert!
TRENDS in Cognitive Sciences

Figure 3. Illustration of failure of auditory streaming. Two brothers address their mother simultaneously. Although the local spectro-temporal structure of the speech signals supports formation of words (local objects), the words are not properly sorted into streams, and she does not properly perceive either message.

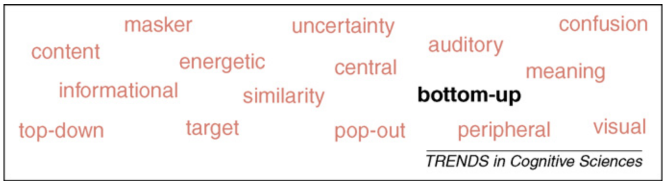

Figure 4. Visual analogy illustrating how object selection can be driven by bottom-up salience. In this example, objects form based primarily on the spatial proximity of the letters within, compared with across, words in the image. Thus, object formation is not an issue; letters form automatically into meaningful words. The phrase 'bottom-up' pops out because it is different from, and more salient than, the other words: attention is automatically drawn to this phrase even in the absence of any top-down desire to attend to it. However, if a viewer is specifically told to look at the bottom left corner of the image, the phrase 'top-down' becomes the focus of attention. In order for volitional attention to override bottom-up salience and select a desired target, the observer must be told some feature (here, spatial location) that differentiates the target from the competing objects.

Failures of object selection

Consistent with the theory of biased competition, volitional selection of an object occurs through top-down attention. If the target object has features that differentiate it from other objects in a scene and if the listener knows these distinctive features *a priori*, s/he can properly direct attention to select the target.

Failures in object selection can occur because a listener directs attention to the wrong object, either because they do not know what feature to attend or because the target and masker features are not sufficiently distinct to ensure proper target selection (attending to the wrong male voice) [39,40]. Indeed, many studies of informational masking using speech signals demonstrate failures of object selection: listeners might perceive a properly formed stream of words (objects form properly), but report a masker rather than the target stream.

Even when the listener is sure of which object is the target, object selection can fail when a competing object is inherently more salient (e.g. much louder) than the target [41]. In these cases, the top-down bias of attention is insufficient to override bottom-up salience and win the biased competition [41]. Figure 4 illustrates the influence of bottom-up salience on attention, again by visual analogy. In the auditory domain, anything from an unexpected, loud sound (a door slamming) to a signal that has special significance (your name being spoken from across the room) can draw attention involuntarily through bottom-up salience [41].

The more unique and distinct the target features, the more effective top-down attention is in enhancing the target and suppressing any maskers [42]. Thus, object selection is a probabilistic competition that depends on interactions between bottom-up and top-down biases [43].

Summary

In both vision and audition, we direct top-down attention to select desired objects from a complex scene. Because perceptual objects are the basic units of attention, proper object formation is crucial to this ability. Stimulus structure determines how objects form locally, either in space-time (for visual objects) or time-frequency (for auditory objects). Higher-order perceptual attributes enable both

object formation across larger scales and selection of a desired object from a complex scene. In complex settings, interactions between object formation and object selection are crucial in enabling us to manage the flow of sensory information we receive. The similarities between auditory and visual perception in complex scenes suggest that common neural mechanisms control attention across modalities. Moreover, a framework based on auditory object formation and auditory object selection can help explain results of many recent psychoacoustic experiments.

Acknowledgements

Grants from NIDCD, AFOSR, ONR and NSF supported this work. These ideas were developed through discussions with Gin Best, Antje Ihlefeld, Erick Gallun, Chris Mason, Gerald Kidd, Steve Colburn and Nat Durlach.

References

1 Simons, D.J. and Rensink, R.A. (2005) Change blindness: past, present, and future. *Trends Cogn. Sci.* 9, 16–20

2 Desimone, R. and Duncan, J. (1995) Neural mechanisms of selective visual attention. *Annu. Rev. Neurosci.* 18, 193–222

3 Serences, J.T. *et al.* (2004) Preparatory activity in visual cortex indexes distractor suppression during covert spatial orienting. *J. Neurophysiol.* 92, 3538–3545

4 Feldman, J. (2003) What is a visual object? *Trends Cogn. Sci.* 7, 252–256

5 Whalen, D.H. and Liberman, A.M. (1996) Limits on phonetic integration in duplex perception. *Percept. Psychophys.* 58, 857–870

6 Darwin, C.J. (1995) Perceiving vowels in the presence of another sound: a quantitative test of the “old-plus-new” heuristic. In *Levels in Speech Communication: Relations and Interactions, A Tribute to Max Wajskop* (Sorin, J.C. *et al.*, eds), pp. 1–12, Elsevier

7 Shinn-Cunningham, B.G. *et al.* (2007) A sound element gets lost in perceptual competition. *Proc. Natl. Acad. Sci. U. S. A.* 104, 12223–12227

8 Cusack, R. *et al.* (2004) Effects of location, frequency region, and time course of selective attention on auditory scene analysis. *J. Exp. Psychol. Hum. Percept. Perform.* 30, 643–656

9 Sussman, E.S. *et al.* (2007) The role of attention in the formation of auditory streams. *Percept. Psychophys.* 69, 136–152

10 Carlyon, R.P. *et al.* (2001) Effects of attention and unilateral neglect on auditory stream segregation. *J. Exp. Psychol. Hum. Percept. Perform.* 27, 115–127

11 Pressnitzer, D. and Hupe, J.M. (2006) Temporal dynamics of auditory and visual bistability reveal common principles of perceptual organization. *Curr. Biol.* 16, 1351–1357

12 Bregman, A.S. (1990) *Auditory Scene Analysis: The Perceptual Organization of Sound,* The MIT Press

13 Darwin, C.J. and Carlyon, R.P. (1995) Auditory grouping. In *Hearing: Handbook of Perception and Cognition* (Moore, B.C.J., ed.), pp. 387–424, Academic Press

14 Carlyon, R.P. (2004) How the brain separates sounds. *Trends Cogn. Sci.* 8, 465–471

15 Sach, A.J. and Bailey, P.J. (2004) Some characteristics of auditory spatial attention revealed using rhythmic masking release. *Percept. Psychophys.* 66, 1379–1387

16 Darwin, C.J. and Hukin, R.W. (2000) Effectiveness of spatial cues, prosody, and talker characteristics in selective attention. *J. Acoust. Soc. Am.* 107, 970–977

17 Duncan, J. (2006) EPS Mid-Career Award 2004: brain mechanisms of attention. *Q. J. Exp. Psychol. (Colchester)* 59, 2–27

18 Shomstein, S. and Yantis, S. (2004) Configural and contextual prioritization in object-based attention. *Psychon. Bull. Rev.* 11, 247–253

19 Yantis, S. (2005) How visual salience wins the battle for awareness. *Nat. Neurosci.* 8, 975–977

20 Serences, J.T. *et al.* (2005) Parietal mechanisms of switching and maintaining attention to locations, objects, and features. In *Neurobiology of Attention* (Itti, L. *et al.*, eds), pp. 35–41, Academic Press

21 Knudsen, E.I. (2007) Fundamental components of attention. *Annu. Rev. Neurosci.* 30, 57–78

22 Busse, L. *et al.* (2005) The spread of attention across modalities and space in a multisensory object. *Proc. Natl. Acad. Sci. U. S. A.* 102, 18751–18756

23 Shomstein, S. and Yantis, S. (2006) Parietal cortex mediates voluntary control of spatial and nonspatial auditory attention. *J. Neurosci.* 26, 435–439

24 Serences, J.T. *et al.* (2005) Coordination of voluntary and stimulus-driven attentional control in human cortex. *Psychol. Sci.* 16, 114–122

25 Alain, C. and Arnott, S.R. (2000) Selectively attending to auditory objects. *Front. Biosci.* 5, D202–D212

26 Scholl, B.J. (2001) Objects and attention: the state of the art. *Cognition* 80, 1–46

27 Best, V. *et al.* (2007) Visually guided attention enhances target identification in a complex auditory scene. *J. Assoc. Res. Otolaryngol.* 8, 294–304

28 Vogel, E.K. and Luck, S.J. (2002) Delayed working memory consolidation during the attentional blink. *Psychon. Bull. Rev.* 9, 739–743

29 Cusack, R. *et al.* (2000) Neglect between but not within auditory objects. *J. Cogn. Neurosci.* 12, 1056–1065

30 Durlach, N.I. *et al.* (2003) Note on informational masking. *J. Acoust. Soc. Am.* 113, 2984–2987

31 Cooke, M. (2006) A glimpsing model of speech perception in noise. *J. Acoust. Soc. Am.* 119, 1562–1573

32 Ciocca, V. and Bregman, A.S. (1987) Perceived continuity of gliding and steady-state tones through interrupting noise. *Percept. Psychophys.* 42, 476–484

33 Warren, R.M. *et al.* (1972) Auditory induction: perceptual synthesis of absent sounds. *Science* 176, 1149–1151

34 Shinn-Cunningham, B.G. and Wang, D. (2008) Influences of auditory object formation on phonemic restoration. *J. Acoust. Soc. Am.* 123, 295–301

35 Kidd, G., Jr *et al.* (2002) Similarity, uncertainty, and masking in the identification of nonspeech auditory patterns. *J. Acoust. Soc. Am.* 111, 1367–1376

36 Best, V. *et al.* (2007) Binaural interference and auditory grouping. *J. Acoust. Soc. Am.* 121, 420–432

37 Culling, J.F. and Summerfield, Q. (1995) Perceptual separation of concurrent speech sounds: absence of across-frequency grouping by common interaural delay. *J. Acoust. Soc. Am.* 98, 785–797

38 Kidd, G., Jr *et al.* (2003) Multiple bursts, multiple looks, and stream coherence in the release from informational masking. *J. Acoust. Soc. Am.* 114, 2835–2845

39 Kidd, G., Jr *et al.* (2005) The advantage of knowing where to listen. *J. Acoust. Soc. Am.* 118, 3804–3815

40 Darwin, C.J. *et al.* (2003) Effects of fundamental frequency and vocal-tract length changes on attention to one of two simultaneous talkers. *J. Acoust. Soc. Am.* 114, 2913–2922

41 Conway, A.R. *et al.* (2001) The cocktail party phenomenon revisited: the importance of working memory capacity. *Psychon. Bull. Rev.* 8, 331–335

42 Durlach, N.I. *et al.* (2003) Informational masking: counteracting the effects of stimulus uncertainty by decreasing target-masker similarity. *J. Acoust. Soc. Am.* 114, 368–379

43 Buschman, T.J. and Miller, E.K. (2007) Top-down versus bottom-up control of attention in the prefrontal and posterior parietal cortices. *Science* 315, 1860–1862

44 Cusack, R. and Carlyon, R.P. (2003) Perceptual asymmetries in audition. *J. Exp. Psychol. Hum. Percept. Perform.* 29, 713–725

45 Winkowski, D.E. and Knudsen, E.I. (2006) Top-down gain control of the auditory space map by gaze control circuitry in the barn owl. *Nature* 439, 336–339

46 Best, V. *et al.* (2006) The influence of spatial separation on divided listening. *J. Acoust. Soc. Am.* 120, 1506–1516

47 Warren, R.M. (1970) Perceptual restoration of missing speech sounds. *Science* 167, 392–393

48 Warren, R.M. *et al.* (1994) Auditory induction: reciprocal changes in alternating sounds. *Percept. Psychophys.* 55, 313–322

49 Best, V. *et al.* (2005) Spatial unmasking of birdsong in human listeners: Energetic and informational factors. *J. Acoust. Soc. Am.* 118, 3766–3773