

Selective Attention in Normal and Impaired Hearing

Barbara G. Shinn-Cunningham, PhD, and Virginia Best, PhD

A common complaint among listeners with hearing loss (HL) is that they have difficulty communicating in common social settings. This article reviews how normal-hearing listeners cope in such settings, especially how they focus attention on a source of interest. Results of experiments with normal-hearing listeners suggest that the ability to selectively attend depends on the ability to analyze the acoustic scene and to form perceptual auditory objects properly. Unfortunately, sound features important for auditory object formation may not be robustly encoded in the auditory periphery of HL listeners. In turn, impaired auditory object formation may interfere with the ability to filter out competing sound sources. Peripheral degradations are also likely to reduce the salience of higher-order auditory cues such as location, pitch, and timbre, which enable normal-hearing listeners to select a desired sound

source out of a sound mixture. Degraded peripheral processing is also likely to increase the time required to form auditory objects and focus selective attention so that listeners with HL lose the ability to switch attention rapidly (a skill that is particularly important when trying to participate in a lively conversation). Finally, peripheral deficits may interfere with strategies that normal-hearing listeners employ in complex acoustic settings, including the use of memory to fill in bits of the conversation that are missed. Thus, peripheral hearing deficits are likely to cause a number of inter-related problems that challenge the ability of HL listeners to communicate in social settings requiring selective attention.

Keywords: attention; segregation; auditory object; auditory scene analysis

Imagine yourself at a restaurant with a group of friends. Conversation trades off from one talker to another. Especially when the topic under discussion is interesting and emotions are high, interruptions are common. Quips and gentle barbs punctuate the conversation, short bursts of levity that add to the feeling of camaraderie. Topics change quickly as one anecdote reminds another talker of some vaguely related idea. In the background, laughter and conversation from nearby tables swirls by.

From Hearing Research Center, Departments of Cognitive and Neural Systems and Biomedical Engineering, Boston University, Boston, and Speech and Hearing Bioscience and Technology Program, Harvard-MIT Division of Health Sciences and Technology, Cambridge, Massachusetts (BGS); Discipline of Physiology, School of Medical Sciences, University of Sydney, Sydney, Australia (VB).

This work was supported by grants from the Office of Naval Research (N000140410131) and the National Institutes of Health (R01 DC05778). Virginia Best was also supported by a University of Sydney Postdoctoral Research Fellowship.

Address correspondence to: Barbara G. Shinn-Cunningham, PhD, 677 Beacon St., Boston, MA 02421; e-mail: shinn@cns.bu.edu.

Most young, normal-hearing (NH) listeners find such settings engaging and exciting. However, for the listener with hearing loss (HL), such a scene can be intimidating and overwhelming (Noble, 2006). Competing sounds can mask other sounds acoustically, rendering parts undetectable. Multiple sources vie for attention at any given moment. The source that is the desired focus of attention shifts suddenly and unpredictably as the conversation evolves. Rapid changes in topic reduce the contextual cues that can help disambiguate the meaning of noisy or partially masked speech.

In such settings, hearing aids can help, especially bilateral aids; however, many listeners are still frustrated and unable to participate in the social interaction, which may result in social isolation (Gatehouse & Akeroyd, 2006; Noble, 2006). One factor that undoubtedly contributes to the problems that HL listeners experience even with amplification is that they have poor frequency resolution (Moore, 2007). As a result, more of a desired signal will be inaudible or distorted. In

addition, HL listeners appear to have a more fundamental problem: They generally have difficulty focusing on one sound source and filtering out unwanted sources (Gatehouse & Akeroyd, 2006).

To understand why HL listeners have difficulty focusing selective attention, we must first understand the processes allowing NH listeners to direct attention to a desired source and comprehend it. Although not intended as an exhaustive literature review, this discussion provides examples, many from the recent literature, that explore the factors allowing NH listeners to communicate in common social settings and that give insight into why peripheral hearing impairments may interfere with everyday communication in common social settings. From this narrative emerges the idea that deficits at early stages of auditory processing can lead to failures of high-level perception because of the way in which different stages of processing build on one another.

Selective Attention in Normal-Hearing Listeners

Perceptual Objects and Attention

Recent work suggests that the same basic principles govern visual and auditory attention (Fritz, Elhilali, David, & Shamma, 2007; Knudsen, 2007; Shamma, 2008; Shinn-Cunningham, 2008). Thus, to build insight into how attention operates to select an acoustic target from a complex auditory scene, here we consider the general mechanisms that govern attention, using evidence from both modalities.

Much of the work on attention builds on the concept of perceptual “objects.” However, it is hard to define precisely what a perceptual object is. Despite the fact that it is difficult to come up with a clear, unambiguous definition, most people have an intuitive understanding of what constitutes an object: the melody of a flute, the flush of a toilet, or the crash of a breaking mirror in an auditory scene; anything from a book lying on a counter to a person’s shadow in a visual scene. Throughout this review, we will adopt a working definition of a perceptual object as a *perceptual estimate* of the sensory inputs that are coming from a distinct physical item in the external world (Shinn-Cunningham, 2008; see also Alain & Tremblay, 2007).

In understanding visual attention, objects are thought to be important because attention operates as a

“biased competition” between the neural representations of perceptual objects (see, e.g., Desimone & Duncan, 1995; Kastner & Ungerleider, 2000). At any one time, one visual object is the focus of attention and is processed in greater detail than other visual objects. Which visual object is in the perceptual foreground depends on an interaction between the inherent salience of the objects competing for attention (a function of their brightness, size, and other attributes) and the goals of the observer (e.g., what color to attend). As in vision (see, e.g., Serences, Shomstein, et al., 2005; Shamma, 2008; Shinn-Cunningham, 2008), evidence supports the idea that what auditory object is the focus of attention depends both on the inherent salience of the sound sources in the environment (e.g., what is loudest or has special behavioral relevance, such as one’s own name; see, e.g., Conway, Cowan, & Bunting, 2001; Moray, 1959; Wood & Cowan, 1995) as well as the top-down goals of the listener (“I want to listen to the source to my right”; see, e.g., Best, Ozmeral, & Shinn-Cunningham, 2007; Fritz et al., 2007; Ihlefeld & Shinn-Cunningham, in press; Kidd, Arbogast, Mason, & Gallun, 2005; Maddox, Alvarez, Streeter, & Shinn-Cunningham, 2008). When observers know that the object they wish to attend has some desired feature (shape, color, pitch, location, timbre, etc.), the effect is to enhance the neural representation of objects that have that feature, biasing the interobject competition to favor those objects (Buschman & Miller, 2007; Elhilali, Fritz, Chi, & Shamma, 2007; Fritz et al., 2007).

Figure 1A illustrates these principles using a visual example. Because the word “involuntary” is different from and more intense than all the other objects in the figure (here, words), it is more salient than any of the other objects, and attention is automatically drawn to it. (Figure 1B is discussed below.) However, if you are instructed that the target word is in the bottom right corner of the figure, you can easily direct attention to that location and process the word “voluntary.”

In both vision and audition, attention seems to operate on objects; therefore, the way in which objects are formed directly impacts how effective you will be when selectively attending to a desired element in a complex scene (e.g., see reviews by Desimone & Duncan, 1995; Knudsen, 2007; Shamma, 2008; Shinn-Cunningham, 2008). For instance, when trying to selectively attend to someone at a crowded reception, how well you are able to perceptually segregate his/her voice from the sound mixture will help determine how effectively you can

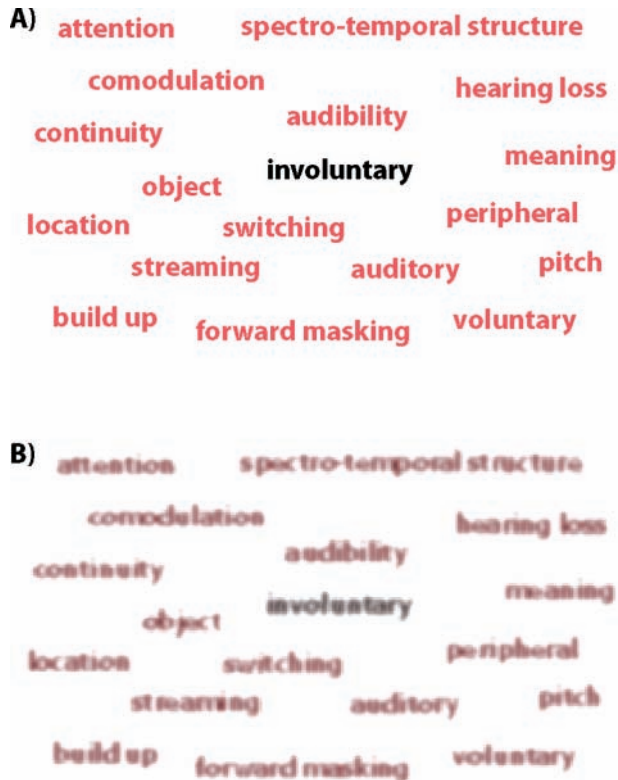


Figure 1. Illustration of how involuntary attention and involuntary attention operate. (A) When the peripheral representation of the words is clear, the word “involuntary” automatically draws attention because it is distinct from the other sources in its color/intensity. However, if attention is directed to the bottom right corner of the panel, the word “voluntary” can be extracted easily. (B) If the peripheral representation is less clear and the colors/intensities of the competing objects are less distinct (as in a listener with hearing loss), involuntary attention is weaker and analysis of each word is more difficult.

tune out the surrounding chatter from other people in the room.

Auditory Object Formation

Given that object formation directly affects the efficacy of selective attention, it is important to consider how auditory objects are formed. Object formation depends on many things, from low-level stimulus attributes to familiarity and expectation (see, e.g., Bregman, 1990). Moreover, the relative importance of different sound attributes for object formation depends on the time scale of analysis (see, e.g., Darwin & Carlyon, 1995).

For short sound elements that are continuous in time and in frequency (e.g., speech vowels or diphthongs), it is the local spectrotemporal structure that most strongly influences object formation (for reviews, see Bregman, 1990; Darwin, 1997; Darwin & Carlyon, 1995). Sound elements with common onsets and common amplitude modulation tend to be perceived as coming from the same source (see, e.g., Best, Gallun, Carlile, & Shinn-Cunningham, 2007a; Culling & Summerfield, 1995). Sounds that are harmonically related tend to group together, as do sounds that are continuous in time–frequency (for reviews, see Bregman, 1990; Carlyon, 2004; Darwin & Carlyon, 1995). On this local time scale, some auditory grouping cues play a relatively weak role. For instance, spatial cues in sound do not have a particularly strong influence on how sound is grouped into objects at this scale of analysis, even though they have some influence (see, e.g., Culling & Summerfield, 1995; Darwin, 2006; Darwin & Ciocca, 1992; Darwin & Hukin, 2000a, 2000b; Drennan, Gatehouse, & Lever, 2003; Shinn-Cunningham, Lee, & Oxenham, 2007).

Although local spectrotemporal structure alone goes a long way toward forming auditory objects, sound coming from a single source often fluctuates over time and has discontinuities and momentary silences. Indeed, spectrotemporal fluctuations are what convey meaning in speech. However, these discontinuities do not typically cause object formation to break down. For instance, the noise-like fricative sound of an “s” is very dissimilar from the spectrotemporal structure of a voiced, pseudoperiodic, and continuous syllable like “no.” However, if a talker utters the word “snow,” the word is usually perceived as one unit. Similarly, during many speech utterances, there are moments of time in which the sound energy dips to zero, such as during the “t” in the middle of a fully articulated utterance of the phrase “fish tank,” yet listeners typically do not have trouble grouping together the phonemes that make up this phrase.

To determine what sounds should group together across spectrotemporal discontinuities, higher-order perceptual features play an important role. These higher-order features include perceived location, pitch, timbre (see, e.g., Culling & Darwin, 1993; Darwin & Hukin, 2000a, 2000b; Kidd et al., 2005), and even signal structure that is learned through experience (e.g., the phonetic, semantic, and lexical structure of speech and language; see, e.g., Bregman, 1990; Warren, 1970). Grouping of temporally disjoint elements of sound (like syllables)

across time is often referred to as *streaming*, and the resulting compound sound is commonly called a *stream* (see, e.g., Bregman, 1990; Shamma, 2008).

Many acoustic mixtures lead to the formation of stable, unambiguous objects. For instance, a serenading violin interrupted by a slamming door is rarely perceived as anything other than two distinct objects. However, in some conditions, the cues determining how objects are formed are contradictory or ambiguous. For instance, if two sound sources happen to turn on and off simultaneously, it is relatively likely that they will be perceived as a single auditory object, even when they actually arise from two distinct sound sources (see, e.g., Best, Gallun, et al., 2007; Bregman, 1990; Woods & Colburn, 1992).

Failures of object formation can impair the ability to analyze a sound source (see, e.g., Best, Gallun, et al., 2007; Darwin & Hukin, 1998; Lee & Shinn-Cunningham, 2008). Specifically, because the meaning of sound is conveyed by its spectrotemporal content, attending to a fusion of multiple sources will interfere with the ability to understand the constituent sources, as the spectrotemporal content of the fused object is not an accurate representation of any single source. Similarly, attending only to a piece of a source (e.g., if a sound is perceived incorrectly as coming from multiple sources rather than as one fused perceptual object) interferes with extracting its meaning.

In many social environments, *words* are formed properly from the mixture of competing speech signals because the spectrotemporal structure of speech provides robust cues for temporally local grouping. However, depending on the setting, it can sometimes be difficult to properly link words together into coherent streams. In particular, if there are multiple talkers in the environment who sound similar, automatic streaming can fail and words from an unwanted talker may intrude and interfere with perception of the desired talker; however, the words themselves are often perceived properly as coherent units that are intelligible (see, e.g., Broadbent, 1958; Brungart, Simpson, Darwin, Arbogast, & Kidd, 2005; Carhart, Tillman, & Greetis, 1969; Darwin & Hukin, 2000a; Ihlefeld & Shinn-Cunningham, 2008a, 2008b; Kidd, Best, & Mason, 2008).

By visual analogy, Figure 2A illustrates the problems that can arise when grouping fails. In the scene, there are a number of letters that can form words. However, because the letters making up the different words are similar in color and size, it is difficult to segregate the letters into words. The immediate, natural way to perceive the visual scene is as one interconnected mass of

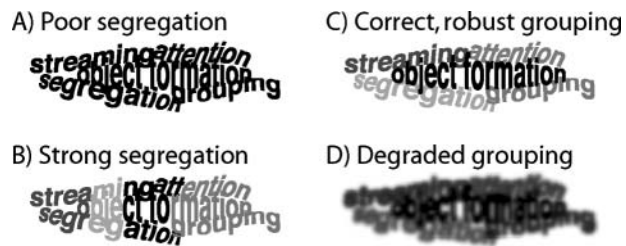


Figure 2. Illustration of how object formation affects selective attention. (A) When all letters are of the same color, the natural way of perceiving the scene is as one object; individual words are difficult to analyze. (B) If colors/intensities of letters differ, different color groups tend to be perceived together. Because the color groups are not word groups, this perceptual organization interferes with processing the individual words. (C) When the colors/intensities of the letters making up each word are distinct, it is easier to attend to each in turn, making understanding more automatic and rapid. (D) If the peripheral representation of the input is less clear and the colors/intensities less distinct from one another (as in a listener with hearing loss), grouping of the letters making up a word is less automatic (as when there are no differences, as in A) and harder to process.

letters. In Figure 2B, different letter groups have different colors/intensities. As a result, the letters tend to group by color. Unfortunately, because the color categories of the letters do not match the word boundaries, cognitive effort is required to link the portions of words that belong together; each word falls in multiple color groups, which interferes with the ability to extract word meaning. Finally, in Figure 2C, the letters of each individual word have a distinctive color/intensity. As a result, letters are grouped properly into words with little or no conscious effort. Understanding each word is much easier: The observer can simply focus attention on each word, one at a time, and process it with little effort.

Auditory Object Selection

Even if objects and streams are formed properly, listeners may not be able to selectively attend to a source of interest. For instance, a listener may hear properly formed words (as a result of short-term object formation) and sentences (as a result of streaming), but still may have trouble selecting which stream to attend (see, e.g., Broadbent, 1952; Brungart, 2001; Brungart, Simpson, Ericson, & Scott, 2001; Ihlefeld & Shinn-Cunningham, 2008a, 2008b; Kidd et al., 2008; Shinn-Cunningham, 2008).

To select a desired stream from a simultaneous mixture, it must have some feature or attribute that

distinguishes it from the other streams in the mixture. Moreover, the listener must have a priori knowledge of what distinguishes the desired stream from the competing streams, or they will be forced to selectively sample each stream in the mixture to test whether it is the desired source. This kind of sampling strategy is known as “serial search” in the vision literature (see, e.g., Wolfe & Horowitz, 2004). During serial search, each additional interfering object in the scene further degrades performance (often measured as an increase in the reaction time to detect a desired visual target).

Serial auditory search arises either because the listener does not know a stimulus feature that distinguishes the object from competing sources (see, e.g., Best, Ozmeral, et al., 2007; Kidd et al., 2005) or because the perceptual features of the objects in the scene are too similar to allow top-down selection to be effective (see, e.g., Brungart et al., 2001). The former kind of failure of object selection could occur in the visual example of Figure 1A if the target word was “voluntary,” but the listener did not know to direct spatial attention to the bottom right of the figure to find the target. An example of the second kind of failure of object selection could arise if you tried to direct attention toward a woman speaking on your right, but there was another woman speaking from nearly the same direction.

If there is a known feature of a visual target that distinguishes it from the other objects in the scene (e.g., the target has a distinct, known color or shape; see, e.g., Wolfe & Horowitz, 2004), adding additional distracting objects does not have a big impact on performance, leading to a far more efficient “parallel search” of the scene. Similarly, there are a number of auditory attributes that listeners can use to direct top-down attention to select a desired object, including source location, pitch, intensity, timbre, spectral content, and rhythm (see, e.g., Best, Ozmeral, et al., 2007; Brungart et al., 2001; Darwin & Hukin, 2000a, 2000b; Kidd et al., 2005; Pitt & Samuel, 1990; Richards & Neff, 2004). However, there may be other features that can guide selective attention to a desired object in an auditory scene; more work is needed to identify the auditory features that can be used in this way.

In an auditory scene, serial search requires first listening to one stream, determining whether or not it is the desired object, and then, if the attending stream is not the desired target of attention, switching attention to another stream. As a result, listeners are very likely to miss some of the content of the stream that they wish to understand during the time that they are

mentally sampling the incorrect streams. Moreover, if a listener must perform a serial search of an auditory scene, he/she will tend to miss more and more of the target message as the number of candidate objects in the scene increases. Thus, the number of objects in and complexity of a scene will affect the speed with which a listener can focus and switch attention to a desired source and, therefore, will affect his/her ability to communicate in social settings with many similar, competing sources.

Build Up of Objects and Switching of Attention

In discussing auditory scene analysis and selective attention, one complication is that the perception of auditory objects can be unstable and labile. Although there are many natural acoustic scenes in which the sources in the scene are sufficiently distinct that the objects and streams form robustly and almost instantaneously, complex scenes often require time for the streaming to “build up.” A well-known demonstration of this is the “ABA” paradigm used in many psychophysical studies (Bregman, 1990; van Noorden, 1975). In this paradigm, perception of a repeating sequence of high-pitched (A), low-pitched (B), and high-pitched (A) tones may initially be perceived as one object (one auditory stream, ABA-ABA). However, perception can change over time until the high-pitched tones are heard in one stream and the low-pitched tones are perceived in another stream (A-A-A-A . . . and -B-B- . . .; see, e.g., Cusack, Deeks, Aikman, & Carlyon, 2004). More generally, it appears that in complex scenes, the way in which the scene is broken down into perceptual objects develops over time as evidence about the spectrotemporal structure of the mixture accrues (Carlyon, 2004; Cusack, Carlyon, & Robertson, 2001; Cusack et al., 2004; Naatanen, Tervaniemi, Sussman, Paavilainen, & Winkler, 2001).

To the extent that the formation of auditory objects builds up over time, selective attention is also likely to become more effective at focusing on a desired source and suppressing competing sources over time. Recent experimental evidence finds that this is indeed the case (Best, Ozmeral, Kopco, & Shinn-Cunningham, 2008; see also Teder-Sälejärvi & Hillyard, 1998, for physiological evidence for such dynamic tuning, albeit to non-speech signals). In a complex speech mixture, the effectiveness of auditory spatial attention improves over time when listeners keep attention focused on a single direction (Best et al., 2008). Moreover, this improvement

in selective attention is enhanced when other nonspatial features (voice quality and temporal continuity) enhance object formation and streaming. This study suggests that time is required to build up object formation and that this build up allows attention to become more selectively focused over time as listeners maintain attention on one object. As the ambiguity in how to organize an acoustic scene decreases, the speed of this build up increases. Conversely, anything that degrades the cues underlying object formation is likely to slow down object formation, which will slow down the refinement of selective attention.

The finding that selective attention improves with time in a complex scene has an important practical implication. In particular, the longer it takes for selective attention to become focused, the more challenging a particular acoustic environment will be for a listener. As already noted, in conversations among more than a pair of talkers the talker of interest changes unpredictably as the conversation evolves. The more people participating in a conversation, the more rapid and unpredictable the required switches of attention will be. Logically, the slower listeners are at refining selective attention to focus on a talker of interest, the less able they will be to follow a lively conversation.

Coping With Failures of Selective Attention

Listeners often miss portions of a desired stream. Even when object formation and selection are working well, some portion of the stream is likely to be inaudible in those frequencies and at those moments when interfering streams are very intense. Momentary lapses in object formation and object selection, as well as time lost in switching attention to a new object of interest, will also result in a listener perceiving only portions of a desired stream. However, listeners can often understand the message of interest, even when they hear only glimpses of the stream they want to attend (see, e.g., Cooke, 2006; Miller & Licklider, 1950; Warren, 1970).

When the stream of interest is a nonspeech signal, this perceptual filling in is based primarily on spectrotemporal continuity (a process known as “auditory induction”; Petkov, O’Connor, & Sutter, 2003; Warren, Bashford, Healy, & Brubaker, 1994). When the stream of interest is speech, this automatic filling in is known as “phonemic restoration,” and the missing speech signal is filled in using not only knowledge of the basic spectrotemporal continuity of the signal but also expectation based on the phonetic, semantic, and linguistic

content of the glimpses (Bashford, Meyers, Brubaker, & Warren, 1988; Bashford & Warren, 1987; Samuel, 2001; Shinn-Cunningham & Wang, 2008; Warren, Hainsworth, Brubaker, Bashford, & Healy, 1997).

Because perceptual filling in makes use of context and word meaning, it will be less effective when the predictability of a message is low. For instance, imagine hearing the phrase “He threw the dog his one” (where one denotes a missed bit of the utterance). The local context will cause most people to guess that the incomplete word is “bone.” However, if the preceding conversation was about why a toddler is crying, even though he was just given an ice cream cone, most listeners would perceive the missing word as “cone,” and often would not even realize that the actual speech signal that they heard was ambiguous. Practically speaking, then, phonemic restoration will be less effective when the topic of conversation changes rapidly and unpredictably and context is less predictable, as when there are many people in a conversation.

When a portion of a message is missed, listeners can also mentally replay sounds they hear from memory. Indeed, evidence supports the idea that when listeners are trying to monitor multiple simultaneous messages (e.g., in a divided attention task), they employ exactly this mechanism, listening actively to one stream and then recalling the other from a short-term store (see, e.g., Best, Gallun, Ihlefeld, & Shinn-Cunningham, 2006; Broadbent, 1958; Ihlefeld & Shinn-Cunningham, 2008a; Pashler, 1998). Although memory can help fill in missing bits of a desired stream, it is volatile and degrades with time (see, e.g., Braida et al., 1984; Brown, 1958; Kidd, Mason, & Hanna, 1988), limiting the circumstances under which such recall is useful. This suggests that if the sensory input to the volatile memory trace is already degraded (e.g., because of external noise), the stored information may not be useful by the time a listener tries to recall it. Recent evidence supports the idea that peripheral, sensory degradations reduce the effectiveness of recalling an acoustic input. Specifically, when listeners are required to divide attention between two simultaneous signals, they appear to actively attend one (the “high-priority” signal) and then recall the other from memory (see, e.g., Ihlefeld & Shinn-Cunningham, 2008a). Degradation of the peripheral representation of competing sources has only a modest impact on the high-priority signal that is actively attended, but causes a disproportionately large degradation of the intelligibility of the lower priority object (Best, Ihlefeld, Mason, Kidd, & Shinn-Cunningham, 2007b).

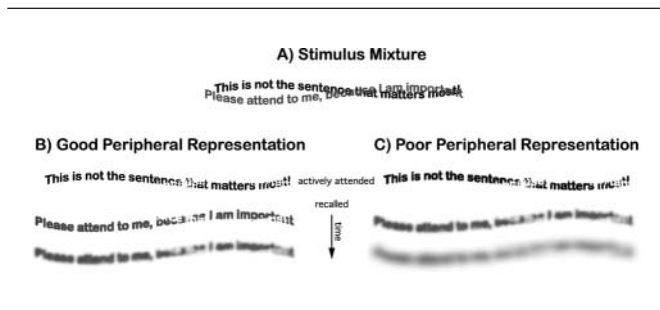


Figure 3. Illustration of how filling in and memory recall can help a listener hearing two competing utterances. In both cases, the listener actively attends the darker message starting with “This” and then tries to recall the other message (starting with “Please”). (A) Stimulus input is two overlapping messages. (B) When the peripheral representation is robust, the actively attended message and the recalled message may both be intelligible, but the recalled message quality degrades rapidly with time. (C) If the peripheral representation is degraded (as in a listener with hearing loss), the actively attended message may be intelligible, but the recalled message may be too degraded to be intelligible even if the listener tries to recall it rapidly.

Figures 3A and B illustrate these ideas. Imagine that a listener initially attends to the stream beginning with the word “This” from a two-stream mixture. The result will be that the appropriate sentence is brought to the perceptual foreground for analysis. Although portions of the stream are inaudible (visually masked in Figure 3B), there may well be enough information to extract the meaning of the utterance, just as in the visual cartoon. If, at some point, listeners realize they need to process the other utterance, they can try to recall the stream beginning with the word “Please,” but the representation of this recalled stream will be degraded, and this degradation will increase with time. Thus, intelligibility of the second stream will generally be poorer than the initially attended stream, and unintelligible if too much time passes before it is recalled from the volatile sensory store.

Automatic perceptual filling in and recall from memory can help listeners determine the content of a desired stream when they fail to selectively attend in a complex scene. However, both of these mechanisms require processing resources and, therefore, add to the cognitive load and to the time it takes for a listener to process and understand a message (see, e.g., Pichora-Fuller, Schneider, & Daneman, 1995). The more challenging the acoustic setting, the harder it is to process a desired signal fast enough to keep up with the flow of information, a critical factor in everyday

communication. In particular, as the number of talkers in and complexity of an acoustic scene increase, more and more of a desired target source will become inaudible because of mutual masking of the competing sources, object formation and object selection will become more challenging, and switches of attention will be required more often. All these factors conspire to increase the effort required to maintain communication. Considered in this light, the ability of young, NH listeners to communicate in a crowded bar is an amazing feat. However, this realization also makes clear why even modest peripheral hearing impairments can profoundly affect communication in HL listeners.

Selective Attention in Listeners With Hearing Loss

Impairments in Auditory Object Formation

To be efficient at selectively attending, listeners must be able to enhance the representation of a source of interest. Simultaneously, they must suppress sources that are not the focus of attention, yet still maintain some awareness of them (to enable rapid refocusing of attention when necessary). To achieve these goals, auditory object formation must be robust. Unfortunately, many of the cues that enable object formation are degraded in HL listeners. This kind of difficulty may help explain why listening selectively in complex settings is particularly challenging for HL listeners.

HL listeners have reduced temporal and spectral acuity when compared with NH listeners (see, e.g., Bernstein & Oxenham, 2006; Carlyon, Long, Deeks, & McKay, 2007; Deeks & Carlyon, 2004; Gatehouse, Naylor, & Elberling, 2003; Leek & Summers, 2001). Broader than normal frequency selectivity in HL listeners results in fewer independent frequency channels representing the auditory scene, making it harder to perceptually segregate the component sources (see, e.g., Gaudrain, Grimault, Healy, & Bera, 2007). In addition, the onsets, offsets, modulation, and harmonic structure important for forming objects over short time scales (e.g., for forming syllables from a sound mixture composed of many talkers) seem to be less perceptually distinct for HL listeners than NH listeners (see, e.g., Bernstein & Oxenham, 2006; Buss, Hall, & Grose, 2004; Hall & Grose, 1994; Kidd, Arbogast,

Mason, & Walsh, 2002; Leek & Summers, 2001; Moore, Glasberg, & Hopkins, 2006).

HL listeners also appear to have difficulty encoding the spectrotemporal fine structure in sounds. Growing evidence supports the idea that spectrotemporal fine structure is critical for robust pitch perception, for speech intelligibility in noise, and for the ability to make effective use of target object information in moments during which an interfering source is relatively quiet (known as “listening in the dips”; see, e.g., Lorenzi, 2008; Pichora-Fuller, Schneider, MacDonald, Brown, & Pass, 2007; Rosen, 1992). In terms of object formation, fine structure may enable a listener to segregate target energy from masker energy (or recognize target epochs) and form a coherent object from the discontinuous target glimpses. HL listeners are inefficient at listening in the dips of a modulated masker (see, e.g., Bronkhorst & Plomp, 1992; Duquesnoy, 1983; Festen & Plomp, 1990; Hopkins, Moore, & Stone, 2008; Lorenzi, Gilbert, Carn, Garnier, & Moore, 2006), probably because their auditory periphery fails to encode spectrotemporal fine structure robustly (see, e.g., Buss et al., 2004; Moore et al., 2006).

On a related note, NH listeners show reduced perceptual interference when spectral bands of interfering sounds are modulated by the same envelope (Hall & Grose, 1994), an effect known as “comodulation masking release.” Comodulation masking release is thought to improve the perceptual segregation of target and interfering sounds because the competing noise bands have correlated envelopes, leading to improvements in target perception. However, HL listeners often show less release from perceptual interference when competing sounds share common modulation, distinct from that of the target (see, e.g., Hall, Davis, Haggard, & Pillsbury, 1988; Hall & Grose, 1994; Moore, Shailer, Hall, & Schooneveldt, 1993).

Both of the above examples are consistent with the idea that the basic spectrotemporal structure of sound, which is critical for grouping together energy from the same source and segregating energy from competing sources, is poorly represented in the auditory system of HL listeners. This failure of object formation will reduce the efficacy of biased competition between objects, which can suppress objects outside the focus of attention (see, e.g., Desimone & Duncan, 1995).

Figure 2D visually illustrates how degradations in the sensory representation of objects in a scene can interfere with object formation and object understanding. In a visual scene, boundaries and edges are important features determining object identity and meaning.

In Figure 2D, edges are blurred, an effect that is somewhat analogous to the poor spectrotemporal resolution typically found in HL listeners. This blurring interferes with the ability to segregate letters from one another. Moreover, other features that can help in grouping and streaming (color in the visual analogy) are also less distinct, further interfering with grouping and understanding. For instance, in Figure 2D, each word can be segregated from the others, but because of the similarity of the colors of the letters making up each word, more effort is required to segregate and analyze a given word than when the letters are clearer and the colors of each word more distinct (Figure 2C).

The perceptual effects of peripheral degradations in the auditory system may be relatively modest when listening in quiet (and easily addressed by simple amplification). However, peripheral degradations are likely to interfere with and slow down object formation. Although only a few studies hint that there are deficits in auditory object formation in HL listeners (for examples, see Healy & Bacon, 2002; Healy, Kannabiran, & Bacon, 2005; Turner, Chi, & Flock, 1999), many studies have not stressed the listener by requiring them to keep up with the processing of an ongoing stream of information (like the situation faced when listening to a conversation at a cocktail party). Increasing the cognitive load by putting listeners in unpredictable settings with ongoing sources could reveal impairments in object formation not observed when listeners perform relatively simple tasks (e.g., using short target utterances may allow listeners to compensate for their deficits and slower processing by “catching up” during the pauses between trials). It has also been suggested that physiological recordings (such as event-related potentials), which can provide highly sensitive measures of object formation, may be useful in future attempts to examine object formation in HL listeners (Alain & Tremblay, 2007).

Impairments in Auditory Object Selection

The same loss of spectrotemporal detail in the periphery that may interfere with object formation is also likely to “muddy” perception of the higher-order features that distinguish a source of interest from interferers and enable selection of the proper focus of attention. In other words, a degraded representation of the auditory scene can result in a target object that is perceptually similar to competing objects. If this occurs, then top-down attention will not be very selective when determining what sound

is perceptually enhanced and what sound is suppressed; imprecise object formation will lead to imprecise object selection.

There are numerous demonstrations that hearing impairment interferes with object selection. HL listeners benefit less from differences in spatial location than NH listeners when trying to follow one talker in the presence of others (see, e.g., Arbogast, Mason, & Kidd, 2005; Bronkhorst & Plomp, 1992; Marrone, Mason, & Kidd, 2007). Recent data show that this is due to both a reduction in the salience of the target as well as a reduced ability to selectively enhance the target based on spatial location (Best, Marrone, Mason, Kidd, & Shinn-Cunningham, 2007). In addition, HL listeners are poorer than NH listeners at hearing out a melody from a mixture of competing melodies and show impaired stream segregation based on voice characteristics, presumably because of reduced spectral resolution in the auditory periphery (see, e.g., Gaudrain et al., 2007; Grose & Hall, 1996; Mackersie, Boothroyd, & Prida, 2000; Mackersie, Prida, & Stiles, 2001).

Figures 1B and 2D demonstrate these concepts through visual analogy. In Figure 1B, the words group properly based on proximity of the letters making up each word compared with the spacing between words. However, because the colors of each word are more similar than in Figure 1A, the word “involuntary” is much less inherently salient than when its color is more distinct. Similarly, in Figure 2D, color similarity conspires with the fuzzy, degraded representation of each word to make it hard to selectively separate one word from the other words. Whereas in Figure 2C, directing attention to the black letters makes it easy to read “object formation,” directing attention to the black letters in Figure 2D is less effective at isolating the desired letters. In the degraded representation, extra time is required to pull the phrase out of the mixture, both because the letters in distracting words are perceptually closer to black and because the phrase itself is not well defined as an object.

Impairments in Object Build Up and Switching Attention

Consistent with the visual analogy described above, degraded peripheral processing is likely to increase the time required to form auditory objects. Given that the focus of selective attention improves as object formation evolves (Best et al., 2008), slowing of object formation will impede the build up of selective attention.

A slowing of selective processing is likely to cause a listener to miss portions of a sound source of interest as he/she struggles to resolve the desired source from the competition. This loss is likely to be particularly problematic when attention must switch rapidly between objects, such as in lively and dynamic conversations. Specifically, each shift of attention is likely to reset object formation and slow object selection (see, e.g., Best et al., 2008; Cusack et al., 2004; Macken, Tremblay, Houghton, Nicholls, & Jones, 2003). As a result, an HL listener is likely to have difficulty keeping up with the flow of information when attention must constantly be redirected. Although very few studies have used dynamic listening situations to demonstrate effects of hearing impairment on auditory object build up and switching of attention (but see Gatehouse & Akeroyd, 2008; Singh, Pichora-Fuller, & Schneider, 2008), it is precisely these environments that most strongly evoke feelings of handicap in HL listeners (Gatehouse & Noble, 2004).

Unfortunately, to compound these problems, many HL listeners are also elderly. Aging has been shown to cause general changes in many cognitive processes, including impairment of executive function and deficits in the ability to filter out unwanted distractions (see Treitz, Heyder, & Daum, 2007; Tun, 1998; Tun, O’Kane, & Wingfield, 2002; but see also Li, Daneman, Qi, & Schneider, 2004; Schneider, Li, & Daneman, 2007). Aging also can affect basic auditory abilities such as temporal perception, even when there is little HL (Gordon-Salant, Fitzgibbons, & Friedman, 2007; Gordon-Salant, Yeni-Komshian, Fitzgibbons, & Barrett, 2006; Grose, Hall, & Buss, 2006; Pichora-Fuller, Schneider, Hamstra, Benson, & Storzer, 2006). A large body of research suggests that cognitive declines and perceptual factors interact to make communication difficult in elderly listeners (see, e.g., Craik, 2007; Humes, 2007; Pichora-Fuller, 2003; Pichora-Fuller & Souza, 2003). On the other hand, it has also been argued that many of the cognitive deficits observed with aging are in fact downstream *consequences* of degraded perceptual representations (see, e.g., Pichora-Fuller et al., 1995; Schneider, Daneman, & Murphy, 2005; Schneider, Daneman, & Pichora-Fuller, 2005). Indeed, one theme of this review is that peripheral deficits contribute to failures in later processing stages (including selective attention) because the normal stages of processing build on one another. Thus, poorer spectrotemporal resolution in the auditory periphery will manifest as an inability to selectively

attend, particularly in settings where selective attention must be deployed rapidly and switch often to be effective (see, e.g., Murphy, Daneman, & Schneider, 2006).

Why Coping Mechanisms That Aid Normal-Hearing Listeners Are Not Enough

Higher auditory thresholds and reduced frequency resolution mean that a greater percentage of a target acoustic signal will be inaudible in HL listeners compared to NH listeners. Although perceptual filling in (e.g., auditory induction, phonemic restoration) can allow listeners to make sense of a desired message even when portions of it are inaudible or missed, HL listeners will tend to require more filling in than do NH listeners. Because perceptual filling in drains cognitive resources and increases processing effort, an HL listener will have to work harder to make sense of a signal in a complicated acoustic setting.

When competing sounds fluctuate (such as when they are speech signals, like at a cocktail party), NH listeners are able to extrapolate the meaning based on the small fragments or glimpses of clean speech available in moments that the interferers are relatively quiet (see, e.g., Cooke, 2006). However, HL listeners are less able than NH listeners to make use of target glimpses (see, e.g., Lorenzi, 2008). Thus, HL listeners are in the difficult position of both hearing fewer glimpses of a target signal and being less effective at making use of the glimpses they do hear. Although contextual cues can alleviate these difficulties in many circumstances (Wingfield & Tun, 2007), they can do very little when conversation is unpredictable. Indeed, elderly listeners appear to rely more heavily on context to make sense of speech in complex settings, presumably to compensate for peripheral HL and less efficient selective attention (Divenyi, 2005), and thus may have disproportionate difficulties when context is less predictable.

The fact that object formation and object selection may be slower and less effective in HL listeners than in NH listeners will add to the cognitive load HL listeners experience when attempting to process an object of interest. However, an HL listener and an NH listener may perform similarly on a test of speech intelligibility, even though the HL listener expends greater effort to achieve that performance (McCoy et al., 2005). For speech tasks, increases in

cognitive load can be inferred in various ways (such as impaired performance on a concurrent task) even in cases where intelligibility does not suffer (see, e.g., McCoy et al., 2005). Increased load has been documented in both HL listeners (Rakerd, Seitz, & Whearty, 1996) and NH listeners presented with noise-degraded signals (Broadbent, 1958; Rabbitt, 1966). Moreover, subjective reports of HL listeners often point to how tiring it is to communicate in complex settings, consistent with them working harder than NH listeners in such environments (Gatehouse & Noble, 2004; McCoy et al., 2005).

Listeners with HL may also be unable to rely on immediate auditory memory to recover missing pieces of a stream of information. Because HL leads to longer and more frequent lapses in intelligibility, HL listeners will miss larger portions of a target signal than NH listeners. Unfortunately, because the raw sensory input to memory is degraded in HL listeners, the volatile memory store may not be very helpful. This would also impact heavily on complex tasks such as the processing of simultaneous messages that make use of this temporary store (see, e.g., Mackersie et al., 2000, 2001). This is illustrated in Figure 3C, where degraded sensory inputs may be unintelligible when further degraded by the time spent in the volatile memory store. In addition, there is evidence that degraded peripheral representations and/or the increased processing demands caused by hearing impairment and aging interfere with the long-term storage and long-term recall of a message (McCoy et al., 2005; Pichora-Fuller et al., 1995).

Although HL listeners are able to understand speech in many noisy situations, the factors just described illustrate why this ability is less robust and more challenging than for NH listeners. HL listeners must work harder than NH listeners to segregate, select, clean up, fill in, and store a message of interest when listening in a complex setting (see also Pichora-Fuller, 2007). Settings that pose no challenge to NH listeners may require HL listeners to exert real effort to communicate. In a scene that is modestly challenging for NH listeners, like joining in a spirited discussion at a bar, HL listeners may fail to keep up, hearing only bits and pieces of the conversation that are too isolated to convey meaning.

Current Hearing Aid Approaches

Many hearing aids have been designed to improve the intelligibility of speech (see, e.g., Byrne, Dillon,

Ching, Katsch, & Keidser, 2001; Rankovic, 1991), including algorithms to enhance spectrotemporal contrast with the hope of increasing the amount of information the HL listener can extract about the spectrotemporal structure of the sound (see, e.g., Baer, Moore, & Gatehouse, 1993; Bunnell, 1990). These efforts are laudable: If a listener cannot understand a sound in quiet, he/she will never be able to selectively attend and understand that sound when it is presented in a complex acoustic scene.

However, in many everyday settings, the best approach would be to amplify the current source of interest but *suppress* competing sources, reducing the amount of effort required to selectively attend to the target stream. Two common approaches toward achieving this goal are to implement noise suppression algorithms and to build directional hearing aids that enhance the sound source from in front of the listener (for reviews, see Dillon, 2001; Dillon & Lovegrove, 1993; Levitt, 2001; Ricketts & Dittberner, 2002). Yet neither of these approaches works particularly well if the listener is trying to participate in a multiperson conversation.

Noise suppression algorithms work well when suppressing sources that are relatively stationary (i.e., that have spectral content that is either unchanging or changing slowly over time). Such algorithms are thus useful for suppressing “uninteresting” sounds (such as the noise of the air conditioner), which can have the net effect of making “interesting” sounds (such as speech) relatively more salient and audible. However, in a multiperson conversation, the interruptions that mask the current talker of interest are nonstationary, unpredictable, and statistically similar to the target speech. As a result, noise suppression approaches are not very effective at dealing with these situations.

On the surface, a directional hearing aid embodies many of the attributes needed to improve speech understanding in a multiperson conversation. Most directional hearing aids are set up to automatically suppress sources from directions other than straight ahead, under the reasonable assumption that a listener will face the person whom he/she wishes to attend. This will enhance the speech from straight ahead, making it more salient and easier to selectively pick out of the sound mixture. As a result, a directional hearing aid can help a listener selectively attend to a sound in front of him/her.

Unfortunately, one of the hallmarks of an animated discussion is that the conversation flows from one talker to the next in unpredictable ways. Whenever the

talker changes, the user of a directional hearing aid must detect the change and turn to face the new talker. Wearing a sharply tuned directional hearing aid is equivalent to wearing blinders: The hearing aid will reduce distraction from sources to the side; however, as a direct result of its design, it will also make events from the side less salient and less able to grab the attention of the listener.

Normal selective attention also acts much like a set of acoustic blinders. However, there is a key difference between how NH listeners use selective attention and how a directional hearing aid works. Selective attention is steerable, focusing and refocusing on whatever sound source is of interest at a given moment. This refocusing is rapid and automatic, happening within a few hundred milliseconds (literally in the blink of an eye; see, e.g., Best et al., 2008; Serences, Liu, & Yantis, 2005a). In contrast, a directional aid focuses attention on the direction a listener is facing, with no consideration of the current goal or desired focus of attention of the listener. When the conversation moves to the side, the listener may not be able to detect who the new talker is or where the talker is located, making it impossible to reorient the directional amplification in the proper direction. Moreover, even if the hearing aid user is almost instantaneously able to determine the direction to face, the physical act of turning the head is slower than the time required by an NH listener to switch the spatial focus of attention. As a result, a directional hearing aid can be very effective if a listener wants to focus attention on one source for an extended time, ignoring all distractions (such as when listening to a formal presentation in a crowded auditorium). However, such an approach will make it even harder for a listener to switch attention as needed in common social situations.

Thus, there is no current hearing aid technology that can allow HL listeners to operate as effectively as NH listeners do in a complex acoustic setting (for discussion of these issues, see Edwards, 2007). Although noise suppression and directional aids can help in many settings, they do not restore the functional ability to fluidly focus attention on whatever source is immediately important, an ability that is critical if a listener is to participate in everyday social interactions.

Summary and Future Directions

Normal-hearing listeners are able to direct top-down attention to select desired auditory objects from out of a sound mixture. Because perceptual objects are the

basic units of attention, proper object formation is important for being able to selectively attend. To select a desired object, listeners must know the feature that identifies that object and enables them to focus and maintain attention on the desired object. The ability to switch attention at will is important in many social settings. Listeners often miss bits of an attended message as a result of masking from competing sources as well as lapses in object formation, object selection, and attention switching. However, they are able to cope with incomplete messages by filling in the missing bits from glimpses they do hear and by replaying the message from memory. The speed of each processing stage is important, as listeners must be able to keep up with the flow of information to interact with others in a social setting.

Multiple factors conspire to interfere with the ability of HL listeners to communicate when there are multiple talkers. The spectrotemporal structure of sound determines how objects form. However, spectrotemporal detail is not encoded robustly in HL listeners. This degraded peripheral representation is likely to impair and slow down object formation in HL listeners. Impaired object formation is likely to degrade the ability to filter out unwanted sources, which will in turn interfere with the ability to understand the source that is the desired focus of attention. Features that enable object selection are also less distinct in HL listeners, making it difficult for them to select the desired source from a mixture. Because the processes of selective attention are slower, HL listeners are likely to miss more of a desired message as they try to focus and switch attention in social scenes. As more of a message is missed, additional processing is required to perceptually fill in and replay the missing message to understand it. Moreover, these processes are likely to be less effective than for an NH listener. Aging exacerbates these problems in many HL listeners. The overall effect is that HL listeners have much greater processing demands and at best normal processing capabilities. When demands exceed capacity, the result can be a catastrophic failure to keep up with the flow of information.

Current hearing aids can enhance listening in quiet and improve selective attention in fixed and predictable settings. This is important because anything that increases the speed and ease of object formation or object selection will reduce the processing load on a listener and improve the ability to participate in everyday social settings. However, the peripheral resolution

of an impaired auditory system is limited, making it impossible to restore fully the spectrotemporal structure that enables NH listeners to segregate competing objects and communicate in everyday settings.

Existing algorithms for source separation cannot yet separate sources as well as NH listeners do. However, even if perfect source separation algorithms existed, another challenge remains: How to present the resulting segregated perceptual objects to an HL listener in a natural, useful manner. A revolutionary assistive listening device would use robust source separation algorithms to create auditory objects, then use knowledge of which source is the desired focus of attention at a given moment to determine how to display these results. Such a device would emphasize the desired target of attention while still enabling the listener to access some information about the other sources in the environment, enabling the listener to selectively attend, at will, to different objects in the environment.

References

- Alain, C., & Tremblay, K. (2007). The role of event-related brain potentials in assessing central auditory processing. *Journal of the American Academy of Audiology*, *18*, 573-589.
- Arbogast, T. L., Mason, C. R., & Kidd, G., Jr. (2005). The effect of spatial separation on informational masking of speech in normal-hearing and hearing-impaired listeners. *Journal of the Acoustical Society of America*, *117*, 2169-2180.
- Baer, T., Moore, B. C., & Gatehouse, S. (1993). Spectral contrast enhancement of speech in noise for listeners with sensorineural hearing impairment: Effects on intelligibility, quality, and response times. *Journal of Rehabilitation Research and Development*, *30*, 49-72.
- Bashford, J. A., Jr., Meyers, M. D., Brubaker, B. S., & Warren, R. M. (1988). Illusory continuity of interrupted speech: Speech rate determines durational limits. *Journal of the Acoustical Society of America*, *84*, 1635-1638.
- Bashford, J. A., Jr., & Warren, R. M. (1987). Multiple phonemic restorations follow the rules for auditory induction. *Perception & Psychophysics*, *42*, 114-121.
- Bernstein, J. G., & Oxenham, A. J. (2006). The relationship between frequency selectivity and pitch discrimination: Sensorineural hearing loss. *Journal of the Acoustical Society of America*, *120*, 3929-3945.
- Best, V., Gallun, F. J., Carlile, S., & Shinn-Cunningham, B. G. (2007). Binaural interference and auditory grouping. *Journal of the Acoustical Society of America*, *121*, 420-432.
- Best, V., Gallun, F. J., Ihlefeld, A., & Shinn-Cunningham, B. G. (2006). The influence of spatial separation on divided listening. *Journal of the Acoustical Society of America*, *120*, 1506-1516.

- Best, V., Ihlefeld, A., Mason, C. R., Kidd, G., Jr., & Shinn-Cunningham, B. G. (2007). *Divided listening in auditory displays*. Paper presented at the International Congress on Acoustics, Madrid, Spain.
- Best, V., Marrone, N., Mason, C. R., Kidd, G., Jr., & Shinn-Cunningham, B. G. (2007). *Do hearing-impaired listeners benefit from spatial and temporal cues in a complex auditory scene?* Paper presented at the International Symposium on Auditory and Audiological Research, Helsingor, Denmark.
- Best, V., Ozmeral, E. J., Kopco, N., & Shinn-Cunningham, B. G. (2008). Object continuity enhances selective auditory attention. *Proceedings of the National Academy of Sciences*, *105*, 13174-13178.
- Best, V., Ozmeral, E. J., & Shinn-Cunningham, B. G. (2007). Visually guided attention enhances target identification in a complex auditory scene. *Journal of the Association for Research in Otolaryngology*, *8*, 294-304.
- Braida, L. D., Lim, J. S., Berliner, J. E., Durlach, N. I., Rabinowitz, W. M., & Purks, S. R. (1984). Intensity perception. XIII. Perceptual anchor model of context-coding. *Journal of the Acoustical Society of America*, *76*, 722-731.
- Bregman, A. S. (1990). *Auditory scene analysis: The perceptual organization of sound*. Cambridge: MIT Press.
- Broadbent, D. E. (1952). Failures of attention in selective listening. *Journal of Experimental Psychology*, *44*, 428-433.
- Broadbent, D. E. (1958). *Perception and communication*. London: Pergamon.
- Bronkhorst, A. W., & Plomp, R. (1992). Effect of multiple speechlike maskers on binaural speech recognition in normal and impaired hearing. *Journal of the Acoustical Society of America*, *92*, 3132-3139.
- Brown, J. (1958). Some tests of the decay theory of immediate memory. *Quarterly Journal of Experimental Psychology*, *10*, 12-21.
- 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.
- Brungart, D. S., Simpson, B. D., Darwin, C. J., Arbogast, T. L., & Kidd, G., Jr. (2005). Across-ear interference from parametrically degraded synthetic speech signals in a dichotic cocktail-party listening task. *Journal of the Acoustical Society of America*, *117*, 292-304.
- Brungart, D. S., Simpson, B. D., Ericson, M. A., & Scott, K. R. (2001). Informational and energetic masking effects in the perception of multiple simultaneous talkers. *Journal of the Acoustical Society of America*, *110*, 2527-2538.
- Bunnell, H. T. (1990). On enhancement of spectral contrast in speech for hearing-impaired listeners. *Journal of the Acoustical Society of America*, *88*, 2546-2556.
- Buschman, T. J., & Miller, E. K. (2007). Top-down versus bottom-up control of attention in the prefrontal and posterior parietal cortices. *Science (New York)*, *315*, 1860-1862.
- Buss, E., Hall, J. W., III, & Grose, J. H. (2004). Temporal fine-structure cues to speech and pure tone modulation in observers with sensorineural hearing loss. *Ear and Hearing*, *25*, 242-250.
- Byrne, D., Dillon, H., Ching, T., Katsch, R., & Keidser, G. (2001). NAL-NL1 procedure for fitting nonlinear hearing aids: Characteristics and comparisons with other procedures. *Journal of the American Academy of Audiology*, *12*, 37-51.
- Carhart, R., Tillman, T. W., & Greetis, E. S. (1969). Perceptual masking in multiple sound backgrounds. *Journal of the Acoustical Society of America*, *45*, 694-703.
- Carlyon, R. P. (2004). How the brain separates sounds. *Trends in Cognitive Sciences*, *8*, 465-471.
- Carlyon, R. P., Long, C. J., Deeks, J. M., & McKay, C. M. (2007). Concurrent sound segregation in electric and acoustic hearing. *Journal of the Association for Research in Otolaryngology*, *8*, 119-133.
- 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.
- Cooke, M. (2006). A glimpsing model of speech perception in noise. *Journal of the Acoustical Society of America*, *119*, 1562-1573.
- Craik, F. I. (2007). The role of cognition in age-related hearing loss. *Journal of the American Academy of Audiology*, *18*, 539-547.
- Culling, J. F., & Darwin, C. J. (1993). The role of timbre in the segregation of simultaneous voices with intersecting F0 contours. *Perception & Psychophysics*, *54*, 303-309.
- Culling, J. F., & Summerfield, Q. (1995). Perceptual separation of concurrent speech sounds: Absence of across-frequency grouping by common interaural delay. *Journal of the Acoustical Society of America*, *98*, 785-797.
- Cusack, R., Carlyon, R. P., & Robertson, I. H. (2001). Auditory midline and spatial discrimination in patients with unilateral neglect. *Cortex*, *37*, 706-709.
- 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.
- Darwin, C. J. (1997). Auditory grouping. *Trends in Cognitive Sciences*, *1*, 327-333.
- Darwin, C. J. (2006). Contributions of binaural information to the separation of different sound sources. *International Journal of Audiology*, *45*(Suppl. 1), S20-S24.
- Darwin, C. J., & Carlyon, R. P. (1995). Auditory grouping. In B. C. J. Moore (Ed.), *Hearing* (pp. 387-424). Orlando, FL: Academic Press.
- Darwin, C. J., & Ciocca, V. (1992). Grouping in pitch perception: Effects of onset asynchrony and ear of presentation of a mistuned component. *Journal of the Acoustical Society of America*, *91*, 3381-3390.
- Darwin, C. J., & Hukin, R. W. (1998). Perceptual segregation of a harmonic from a vowel by interaural time difference in conjunction with mistuning and onset

- asynchrony. *Journal of the Acoustical Society of America*, 103, 1080-1084.
- Darwin, C. J., & Hukin, R. W. (2000a). Effectiveness of spatial cues, prosody, and talker characteristics in selective attention. *Journal of the Acoustical Society of America*, 107, 970-977.
- Darwin, C. J., & Hukin, R. W. (2000b). Effects of reverberation on spatial, prosodic, and vocal-tract size cues to selective attention. *Journal of the Acoustical Society of America*, 108, 335-342.
- Deeks, J. M., & Carlyon, R. P. (2004). Simulations of cochlear implant hearing using filtered harmonic complexes: Implications for concurrent sound segregation. *Journal of the Acoustical Society of America*, 115, 1736-1746.
- Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual Review of Neuroscience*, 18, 193-222.
- Dillon, H. (2001). *Hearing aids*. New York: Thieme Medical.
- Dillon, H., & Lovegrove, R. (1993). Single microphone noise reduction systems for hearing aids: A review and an evaluation. In G. A. Studebaker & I. Hochberg (Eds.), *Acoustical factors affecting hearing aid performance* (pp. 353-372). Boston: Allyn & Bacon.
- Divenyi, P. (2005). *Humans glimpse, too, not only machines (hommage a Martin Cooke)*. Paper presented at Forum Acusticum 2005, Budapest, Hungary.
- Drennan, W. R., Gatehouse, S., & Lever, C. (2003). Perceptual segregation of competing speech sounds: The role of spatial location. *Journal of the Acoustical Society of America*, 114, 2178-2189.
- Duquesnoy, A. J. (1983). Effect of a single interfering noise or speech source upon the binaural sentence intelligibility of aged persons. *Journal of the Acoustical Society of America*, 74, 739-743.
- Edwards, B. (2007). The future of hearing aid technology. *Trends in Amplification*, 11, 31-46.
- Elhilali, M., Fritz, J. B., Chi, T. S., & Shamma, S. A. (2007). Auditory cortical receptive fields: Stable entities with plastic abilities. *Journal of Neuroscience*, 27, 10372-10382.
- Festen, J. M., & Plomp, R. (1990). Effects of fluctuating noise and interfering speech on the speech-reception threshold for impaired and normal hearing. *Journal of the Acoustical Society of America*, 88, 1725-1736.
- Fritz, J. B., Elhilali, M., David, S. V., & Shamma, S. A. (2007). Auditory attention—Focusing the searchlight on sound. *Current Opinion in Neurobiology*, 17, 437-455.
- Gatehouse, S., & Akeroyd, M. (2006). Two-eared listening in dynamic situations. *International Journal of Audiology*, 45(Suppl. 1), S120-S124.
- Gatehouse, S., & Akeroyd, M. A. (2008). The effects of cueing temporal and spatial attention on word recognition in a complex listening task in hearing-impaired listeners. *Trends in Amplification*, 12, 145-161.
- Gatehouse, S., Naylor, G., & Elberling, C. (2003). Benefits from hearing aids in relation to the interaction between the user and the environment. *International Journal of Audiology*, 42(Suppl. 1), S77-S85.
- Gatehouse, S., & Noble, W. (2004). The Speech, Spatial and Qualities of Hearing Scale (SSQ). *International Journal of Audiology*, 43, 85-99.
- Gaudrain, E., Grimault, N., Healy, E. W., & Bera, J. C. (2007). Effect of spectral smearing on the perceptual segregation of vowel sequences. *Hearing Research*, 231, 32-41.
- Gordon-Salant, S., Fitzgibbons, P. J., & Friedman, S. A. (2007). Recognition of time-compressed and natural speech with selective temporal enhancements by young and elderly listeners. *Journal of Speech, Language, and Hearing Research*, 50, 1181-1193.
- Gordon-Salant, S., Yeni-Komshian, G. H., Fitzgibbons, P. J., & Barrett, J. (2006). Age-related differences in identification and discrimination of temporal cues in speech segments. *Journal of the Acoustical Society of America*, 119, 2455-2466.
- Grose, J. H., & Hall, J. W., III (1996). Perceptual organization of sequential stimuli in listeners with cochlear hearing loss. *Journal of Speech and Hearing Research*, 39, 1149-1158.
- Grose, J. H., Hall, J. W., III, & Buss, E. (2006). Temporal processing deficits in the pre-senescent auditory system. *Journal of the Acoustical Society of America*, 119, 2305-2315.
- Hall, J. W., III, Davis, A. C., Haggard, M. P., & Pillsbury, H. C. (1988). Spectro-temporal analysis in normal-hearing and cochlear-impaired listeners. *Journal of the Acoustical Society of America*, 84, 1325-1331.
- Hall, J. W., III, & Grose, J. H. (1994). Signal detection in complex comodulated backgrounds by normal-hearing and cochlear-impaired listeners. *Journal of the Acoustical Society of America*, 95, 435-443.
- Healy, E. W., & Bacon, S. P. (2002). Across-frequency comparison of temporal speech information by listeners with normal and impaired hearing. *Journal of Speech, Language, and Hearing Research*, 45, 1262-1275.
- Healy, E. W., Kannabiran, A., & Bacon, S. P. (2005). An across-frequency processing deficit in listeners with hearing impairment is supported by acoustic correlation. *Journal of Speech, Language, and Hearing Research*, 48, 1236-1242.
- Hopkins, K., Moore, B. C., & Stone, M. A. (2008). Effects of moderate cochlear hearing loss on the ability to benefit from temporal fine structure information in speech. *Journal of the Acoustical Society of America*, 123, 1140-1153.
- Humes, L. E. (2007). The contributions of audibility and cognitive factors to the benefit provided by amplified speech to older adults. *Journal of the American Academy of Audiology*, 18, 590-603.
- Ihlefeld, A., & Shinn-Cunningham, B. (2008a). Spatial release from energetic and informational masking in a divided speech identification task. *Journal of the Acoustical Society of America*, 123, 4380-4392.

- Ihlefeld, A., & Shinn-Cunningham, B. (2008b). Spatial release from energetic and informational masking in a selective speech identification task. *Journal of the Acoustical Society of America*, 123, 4369-4379.
- Ihlefeld, A., & Shinn-Cunningham, B. G. (in press). Disentangling the effects of spatial cues on selection and formation of auditory objects. *Journal of the Acoustical Society of America*.
- Kastner, S., & Ungerleider, L. G. (2000). Mechanisms of visual attention in the human cortex. *Annual Review of Neuroscience*, 23, 315-341.
- Kidd, G., Jr., Arbogast, T. L., Mason, C. R., & Gallun, F. J. (2005). The advantage of knowing where to listen. *Journal of the Acoustical Society of America*, 118, 3804-3815.
- Kidd, G., Jr., Arbogast, T. L., Mason, C. R., & Walsh, M. (2002). Informational masking in listeners with sensorineural hearing loss. *Journal of the Association for Research in Otolaryngology*, 3, 107-119.
- Kidd, G., Jr., Best, V., & Mason, C. R. (in press). *Listening to every other word: Examining the strength of linkage variables in forming streams of speech*. Manuscript submitted for publication.
- Kidd, G., Jr., Mason, C. R., & Hanna, T. E. (1988). Evidence for sensory-trace comparisons in spectral shape discrimination. *Journal of the Acoustical Society of America*, 84, 144-149.
- Knudsen, E. I. (2007). Fundamental components of attention. *Annual Review in Neuroscience*, 30, 57-78.
- Lee, A. K., & Shinn-Cunningham, B. G. (2008). Effects of frequency disparities on trading of an ambiguous tone between two competing auditory objects. *Journal of the Acoustical Society of America*, 123, 4340-4351.
- Leek, M. R., & Summers, V. (2001). Pitch strength and pitch dominance of iterated rippled noises in hearing-impaired listeners. *Journal of the Acoustical Society of America*, 109, 2944-2954.
- Levitt, H. (2001). Noise reduction in hearing aids: A review. *Journal of Rehabilitation Research and Development*, 38, 111-121.
- Li, L., Daneman, M., Qi, J., & Schneider, B. (2004). Does the information content of an irrelevant source differentially affect speech recognition in younger and older adults? *Journal of Experimental Psychology: Human Perception and Performance*, 30, 1077-1091.
- Lorenzi, C. (2008). Reception of phonetic features in fluctuating background noise maskers. *Journal of the Acoustical Society of America*, 123, 3931.
- Lorenzi, C., Gilbert, G., Carn, H., Garnier, S., & Moore, B. C. (2006). Speech perception problems of the hearing impaired reflect inability to use temporal fine structure. *Proceedings of the National Academy of Sciences of the United States of America*, 103, 18866-18869.
- 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 and Performance*, 29, 43-51.
- Mackersie, C. L., Boothroyd, A., & Prida, T. (2000). Use of a simultaneous sentence perception test to enhance sensitivity to ease of listening. *Journal of Speech, Language, and Hearing Research*, 43, 675-682.
- Mackersie, C. L., Prida, T. L., & Stiles, D. (2001). The role of sequential stream segregation and frequency selectivity in the perception of simultaneous sentences by listeners with sensorineural hearing loss. *Journal of Speech, Language, and Hearing Research*, 44, 19-28.
- Maddox, R., Alvarez, R., Streeter, T., & Shinn-Cunningham, B. G. (2008). *Top-down attention modulates bottom-up streaming based on pitch and location*. Paper presented at the mid-winter meeting of the Association for Research in Otolaryngology, Phoenix, AZ.
- Marrone, N., Mason, C. R., & Kidd, G., Jr. (2007). *Listening in a multisource environment with and without hearing aids*. Paper presented at the International Symposium on Auditory and Audiological Research, Helsingor, Denmark.
- McCoy, S. L., Tun, P. A., Cox, L. C., Colangelo, M., Stewart, R. A., & Wingfield, A. (2005). Hearing loss and perceptual effort: Downstream effects on older adults' memory for speech. *Quarterly Journal of Experimental Psychology*, 58, 22-33.
- Miller, G. A., & Licklider, J. C. R. (1950). The intelligibility of interrupted speech. *Journal of the Acoustical Society of America*, 22, 167-173.
- Moore, B. C. (2007). *Cochlear hearing loss: Physiological, psychological, and technical issues* (2nd ed.). Chichester, UK: Wiley.
- Moore, B. C., Glasberg, B. R., & Hopkins, K. (2006). Frequency discrimination of complex tones by hearing-impaired subjects: Evidence for loss of ability to use temporal fine structure. *Hearing Research*, 222, 16-27.
- Moore, B. C., Shailer, M. J., Hall, J. W., III, & Schooneveldt, G. P. (1993). Comodulation masking release in subjects with unilateral and bilateral hearing impairment. *Journal of the Acoustical Society of America*, 93, 435-451.
- Moray, N. (1959). Attention in dichotic listening: Affective cues and the influence of instructions. *Quarterly Journal of Experimental Psychology*, 11, 55-60.
- Murphy, D. R., Daneman, M., & Schneider, B. A. (2006). Why do older adults have difficulty following conversations? *Psychology and Aging*, 21, 49-61.
- Naatanen, R., Tervaniemi, M., Sussman, E., Paavilainen, P., & Winkler, I. (2001). "Primitive intelligence" in the auditory cortex. *Trends in Neurosciences*, 24, 283-288.
- Noble, W. (2006). Bilateral hearing aids: A review of self-reports of benefit in comparison with unilateral fitting. *International Journal of Audiology*, 45(Suppl. 1), S63-S71.
- Pashler, H. E. (1998). *The psychology of attention*. Cambridge: MIT Press.
- Petkov, C. I., O'Connor, K. N., & Sutter, M. L. (2003). Illusory sound perception in macaque monkeys. *Journal of Neuroscience*, 23, 9155-9161.

- Pichora-Fuller, M. K. (2003). Cognitive aging and auditory information processing. *International Journal of Audiology*, 42, S26-S32.
- Pichora-Fuller, M. K. (2007). Audition and cognition: What audiologists need to know about listening. In C. Palmer & R. Seewald (Eds.), *Hearing care for adults: An international conference* (pp. 71-85). Warrenville, IL: Phonak.
- Pichora-Fuller, M. K., Schneider, B., & Daneman, M. (1995). How young and old adults listen to and remember speech in noise. *Journal of the Acoustical Society of America*, 97, 593-608.
- Pichora-Fuller, M. K., Schneider, B., Hamstra, S., Benson, N., & Storzer, E. (2006). Effect of age on gap detection in speech and non-speech stimuli varying in marker duration and spectral symmetry. *Journal of the Acoustical Society of America*, 119, 1143-1155.
- Pichora-Fuller, M. K., Schneider, B., MacDonald, E., Brown, S., & Pass, H. (2007). Temporal jitter disrupts speech intelligibility: A simulation of auditory aging. *Hearing Research*, 223, 114-121.
- Pichora-Fuller, M. K., & Souza, P. (2003). Effects of aging on auditory processing of speech. *International Journal of Audiology*, 42, S11-S16.
- Pitt, M. A., & Samuel, A. G. (1990). The use of rhythm in attending to speech. *Journal of Experimental Psychology: Human Perception and Performance*, 16, 564-573.
- Rabbitt, P. M. A. (1966). Recognition: Memory for words correctly heard in noise. *Psychonomic Science*, 6, 383-384.
- Rakerd, B., Seitz, P. F., & Whearty, M. (1996). Assessing the cognitive demands of speech listening for people with hearing losses. *Ear and Hearing*, 17, 97-106.
- Rankovic, C. M. (1991). An application of the articulation index to hearing aid fitting. *Journal of Speech and Hearing Research*, 34, 391-402.
- Richards, V. M., & Neff, D. L. (2004). Cuing effects for informational masking. *Journal of the Acoustical Society of America*, 115, 289-300.
- Ricketts, T. A., & Dittberner, A. B. (2002). Directional amplification for improved signal-to-noise ratio: Strategies, measurements, and limitations. In M. Valente (Ed.), *Hearing aids: Standards, options, and limitations* (pp. 274-346). New York: Thieme.
- Rosen, S. (1992). Temporal information in speech: Acoustic, auditory and linguistic aspects. *Philosophical Transactions of the Royal Society of London B*, 336, 367-373.
- Samuel, A. G. (2001). Knowing a word affects the fundamental perception of the sounds within it. *Psychological Science*, 12, 348-351.
- Schneider, B. A., Daneman, M., & Murphy, D. R. (2005). Speech comprehension difficulties in older adults: Cognitive slowing or age-related changes in hearing? *Psychology and Aging*, 20, 261-271.
- Schneider, B. A., Daneman, M., & Pichora-Fuller, M. K. (2002). Listening in aging adults: From discourse comprehension to psychoacoustics. *Canadian Journal of Experimental Psychology*, 56, 139-152.
- Schneider, B. A., Li, L., & Daneman, M. (2007). How competing speech interferes with speech comprehension in everyday listening situations. *Journal of the American Academy of Audiology*, 18, 559-572.
- Serences, J. T., Liu, T., & Yantis, S. (2005). Parietal mechanisms of switching and maintaining attention to locations, objects, and features. In L. Itti, G. Rees, & J. Tsotsos (Eds.), *Neurobiology of attention* (pp. 35-41). Burlington, MA: Academic Press.
- Serences, J. T., Shomstein, S., Leber, A. B., Golay, X., Egeth, H. E., & Yantis, S. (2005). Coordination of voluntary and stimulus-driven attentional control in human cortex. *Psychological Science*, 16, 114-122.
- Shamma, S. (2008). On the emergence and awareness of auditory objects. *PLoS Biology*, 6, 1141-1143.
- Shinn-Cunningham, B. G. (2008). Object-based auditory and visual attention. *Trends in Cognitive Sciences*, 12, 182-186.
- Shinn-Cunningham, B. G., Lee, A. K., & Oxenham, A. J. (2007). A sound element gets lost in perceptual competition. *Proceedings of the National Academy of Sciences of the United States of America*, 104, 12223-12227.
- Shinn-Cunningham, B. G., & Wang, D. (2008). Influences of auditory object formation on phonemic restoration. *Journal of the Acoustical Society of America*, 123, 295-301.
- Singh, G., Pichora-Fuller, M. K., & Schneider, B. A. (2008). Auditory spatial attention in conditions of real and simulated spatial separation by younger and older adults. *Journal of the Acoustical Society of America*, 124.
- Teder-Sälejärvi, W. A., & Hillyard, S. A. (1998). The gradient of spatial auditory attention in free field: An event-related potential study. *Perception and Psychophysics*, 60, 1228-1242.
- Treitz, F. H., Heyder, K., & Daum, I. (2007). Differential course of executive control changes during normal aging. *Neuropsychology, Development, and Cognition*, 14, 370-393.
- Tun, P. A. (1998). Fast noisy speech: Age differences in processing rapid speech with background noise. *Psychology and Aging*, 13, 424-434.
- Tun, P. A., O'Kane, G., & Wingfield, A. (2002). Distraction by competing speech in young and older adult listeners. *Psychology and Aging*, 17, 453-467.
- Turner, C. W., Chi, S. L., & Flock, S. (1999). Limiting spectral resolution in speech for listeners with sensorineural hearing loss. *Journal of Speech, Language, and Hearing Research*, 42, 773-784.
- van Noorden, L. P. A. S. (1975). *Temporal coherence in the perception of tone sequences*. Eindhoven, the Netherlands: Eindhoven University of Technology.
- Warren, R. M. (1970). Perceptual restoration of missing speech sounds. *Science (New York)*, 167, 392-393.
- Warren, R. M., Bashford, J. A., Jr., Healy, E. W., & Brubaker, B. S. (1994). Auditory induction: Reciprocal changes in alternating sounds. *Perception & Psychophysics*, 55, 313-322.
- Warren, R. M., Hainsworth, K. R., Brubaker, B. S., Bashford, J. A., Jr., & Healy, E. W. (1997). Spectral restoration of speech:

- Intelligibility is increased by inserting noise in spectral gaps. *Perception & Psychophysics*, 59, 275-283.
- Wingfield, A., & Tun, P. (2007). Cognitive supports and cognitive constraints on comprehension of spoken language. *Journal of the American Academy of Audiology*, 18, 548-558.
- Wolfe, J. M., & Horowitz, T. S. (2004). What attributes guide the deployment of visual attention and how do they do it? *Nature Reviews*, 5, 495-501.
- Wood, N., & Cowan, N. (1995). 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, and Cognition*, 21, 255-260.
- Woods, W. S., & Colburn, H. S. (1992). Test of a model of auditory object formation using intensity and interaural time difference discrimination. *Journal of the Acoustical Society of America*, 91, 2894-2902.