5
Auditory Selective Attention

Barbara Shinn-Cunningham and Virginia Best

Historical Context

Some of the earliest work examining human selective attention considered auditory communication signals, focusing on the issue of how we choose what to listen to in a mixture of competing speech sounds (commonly referred to as the "cocktail party problem"; Cherry, 1953). The earliest selective attention models were bottleneck theories, which propose that a filtering step restricts how much sensory information is passed on to a limited-capacity processing system. Broadbent initially proposed a strict filter that rejects unwanted information based on physical parameters such as location, pitch, loudness, and timbre (Broadbent, 1958). In Broadbent's theory, listeners can access only very basic properties about a rejected source and cannot extract detailed information about its content. While Broadbent argued that listeners can direct volitional attention to physical sound features to determine what information is filtered out, he also recognized that certain stimuli may capture attention involuntarily, overriding listener goals; that is, what is attended and what is rejected depends on both volitional attention and inherent stimulus salience.

While many of the basic tenets of Broadbent's filter theory guide modern thinking, subsequent observations prompted modifications of the theory. For example, some listeners notice their own name when it occurs in an unattended channel, suggesting that some information from filtered-out signals "gets through" (e.g., see Moray, 1959; Wood & Cowan, 1995; Colflesh & Conway, 2007). Similarly, even when listeners are not consciously aware of the content of an unattended stimulus, they can show priming effects where the ignored stimuli influence perception of subsequent items, demonstrating that some meaning from to-be-ignored streams is extracted (e.g., see Gray & Wedderburn, 1960; Treisman, 1960; Corteen & Wood, 1972; Mackay, 1973; however, see also Lachter et al., 2004). To account for such effects, Treisman (1960) proposed that the filter attenuates unwanted stimuli rather than filtering them out completely.

In the late 1960s and early 1970s, many seminal studies of selective attention explored auditory perception. Following this early work, though, visual studies

overshadowed work in the auditory domain. Indeed, Treisman herself developed the influential feature integration theory (FIT) to describe attention in the visual domain (Treisman & Gelade, 1980). FIT argued that features of different objects in a visual scene are processed in parallel, but only those features that are at an attended location are bound together (integrated) and processed in detail as one object. Work testing such ideas about visual attention flourished in the 1970s and 1980s, while most auditory behavioral studies focused on how information is coded in the auditory periphery, with little consideration of limits related to attention. While some key electroencephalography studies showed that selective attention modulates the strength of early cortical responses to sound (e.g., see Picton & Hillyard, 1974; Hansen & Hillyard, 1980), these results did not strongly influence psychoacoustic studies at that time. A handful of studies published in the 1990s used cuing to discriminate between endogenous and exogenous components of auditory attention, largely inspired by previous studies in the visual domain (e.g., Spence & Driver, 1994; Mondor & Zatorre, 1995; Quinlan & Bailey, 1995), but these again used relatively simple stimuli and tasks with low attentional demands.

When auditory researchers began once again to explore auditory perception where central bottlenecks rather than sensory limitations determined performance, the work was rarely related to modern theories of attention and memory. Instead, the term "informational masking" (IM) was coined (see Kidd et al., 2008). IM is a catchall phrase, encompassing any perceptual interference between sounds that was not explained by "energetic masking" (EM), where EM was defined as interference explained by masking within the auditory nerve (e.g., see Kidd et al., 1994; Oh & Lutfi, 1998; Freyman et al., 1999; Watson, 2005). That is, IM encompassed any suboptimal processing of information, including loss of information due to attentional filtering. Still, key components of modern theories of selective attention, both in visual and auditory science, trace back to early auditory work by Cherry, Broadbent, Treisman, and their contemporaries.

State-of-the-Art Literature Review

Object-Based Attention and Auditory Streams

Recent work on auditory attention draws heavily on key findings from vision science (e.g., see Spence et al., 2001; Cusack & Carlyon, 2003; Shinn-Cunningham, 2008; Snyder et al., 2012). In vision, attention is argued to operate as a "biased competition" between the neural representations of perceptual objects (e.g., see Desimone & Duncan, 1995; Kastner & Ungerleider, 2001). Like Broadbent's early theories, the biased-competition view argues the focus of attention is determined by the interplay between the salience of stimuli (exogenously guided attention) and observer goals (endogenously guided attention). However, biased competition arises specifically between objects, each of

which is a collection of attributes. At any one time, one object is the focus of attention and is processed in greater detail than other objects in the scene. Recently, evidence for such effects in auditory processing has started to emerge from physiological studies (e.g., Chait et al., 2010; Maddox et al., 2012; Mesgarani & Chang, 2012; Middlebrooks & Bremen, 2013).

If selective auditory attention is object based, it is important to define what constitutes an auditory object. Yet it is challenging to come up with a clear, irrefutable definition. Seminal work by Al Bregman (1990) laid out some of the rules governing the perceptual organization of sound mixtures. Later researchers explored these rules to reveal the sound features that lead to object formation (see reviews by Carlyon, 2004; Griffiths & Warren, 2004). For instance, sounds tend to group together if they turn on and off together (i.e., are comodulated), are harmonically related, and are continuous in time and frequency. Such sound features are relatively "local" in time, operating at the timescale of, for example, a speech syllable. However, we perceive ongoing speech as one stream even though there are silent gaps across which "local" features cannot operate. Grouping across acoustic discontinuities is driven by higher order perceptual features, such as location, pitch, and timbre, as well as signal structure learned through experience (e.g., phonetic, semantic, and lexical structure).

The relationship between object formation and auditory selective attention remains a subject of debate. Some argue that objects form only when a stream (an auditory object extending through time) is attended (e.g., Jones, 1976; Alain & Woods, 1997; Cusack et al., 2004). However, other studies suggest that auditory streams form automatically and preattentively (e.g., Bregman, 1990; Macken et al., 2003; Sussman et al., 2007). Most likely, both automatic and attention-driven processes influence stream formation. In cases where low-level attributes are sufficiently distinct to define a stream unambiguously, it will be segregated from a sound mixture even without attention. However, sound mixtures are often ambiguous, in which case attention to a particular perceptual feature may help "pull out" the stream that is attended (e.g., see Alain et al., 2001; Macken et al., 2003). This view is supported by studies that show that perception of a complex auditory scene is refined through time (e.g., see Carlyon et al., 2003; Teki et al., 2013).

Our own work supports the idea that object formation and attention interact. Specifically, when listeners are instructed to attend to one sound feature (e.g., pitch or timbre) and report a stream of words, they make more errors when a task-irrelevant sound feature changes between words, breaking down the perceived continuity of the stream (Maddox & Shinn-Cunningham, 2012). Conversely, when the talker of a target word stream is consistent, continuity of talker identity enhances performance even when talker identity is task irrelevant, such as when listeners are trying to attend to words from a particular location (Best et al., 2008; Bressler et al., 2014). These results

demonstrate that even irrelevant features influence selective attention performance, presumably by influencing object formation. On the other hand, listeners weight a given acoustic cue more when it is task relevant than when it is task irrelevant (Maddox & Shinn-Cunningham, 2012). Together, these results suggest that attention is object based, but that attention itself affects object formation (Fritz et al., 2007; Shinn-Cunningham, 2008; Shinn-Cunningham & Best, 2008; Shamma et al., 2011).

Although the idea of object-based attention came from vision, there is relatively little discussion of the relationship between visual object formation and visual attention. We speculate that this is because auditory objects emerge only through time. Consider that a static two-dimensional picture of a natural scene generally contains enough information for visual objects to emerge without any further information. In contrast, auditory information is typically conveyed by changes in sounds as a function of time. Similarly, only by analyzing sound's time-frequency content can the features and structure that drive auditory stream formation be extracted. "Local" grouping features emerge over 10s of milliseconds, but higher order features can require on the order of seconds to be perceived (e.g., see Cusack et al., 2004; Chait et al., 2010). Object formation can take time (e.g., see Cusack et al., 2004) and can be unstable (e.g., see Hupe et al., 2008). Thus, in natural auditory scenes, it is nearly impossible to discuss selective attention without simultaneously considering object formation. Current theories of auditory object formation and attention deal directly with the fact that auditory objects emerge through time, and that this process may both influence and be influenced by attention (Elhilali et al., 2009; Shamma et al., 2011).

Dimensions of Auditory Selective Attention

Although the "unit" of auditory attention seems to be an auditory object, listeners can direct top-down selective attention by focusing on different acoustic dimensions, many of which also influence object and stream formation (e.g., frequency, location, pitch, timbre, etc.). Other dimensions are even more basic; perhaps the most fundamental feature, given the tonotopic arrangement of the auditory system, is frequency. Some of the earliest experiments exploring auditory selective attention used the probe-signal method to demonstrate that listeners can focus attention on a certain frequency region, which enhances detection of a quiet tone at or near that frequency (e.g., Greenberg & Larkin, 1968; Scharf et al., 1987). Perceived location is another powerful cue for directing selective auditory attention, improving the ability to extract information about a target stream (e.g., see Arbogast & Kidd, 2000; Kidd et al., 2005). There are also examples demonstrating that attention can be directed to pitch (Maddox & Shinn-Cunningham, 2012), level (e.g., attending to the softer of two voices; Brungart, 2001; Kitterick et al., 2013), and talker characteristics such as timbre and gender (e.g., Culling et al., 2003; Darwin et al., 2003). Auditory attention can also be focused in time, such that sounds occurring at expected times are better detected than those occurring

at unpredictable times (Wright & Fitzgerald, 2004). This idea has been elaborated to describe attention that is distributed in time either to enhance sensitivity to target sequences ("rhythmic attention"; e.g., Jones et al., 1981) or to cancel irrelevant sounds (Devergie et al., 2010).

Evidence from the neuroimaging literature suggests that selective spatial auditory attention engages some of the same frontoparietal circuitry that controls spatial visual attention (e.g., see Tark & Curtis, 2013; Lee et al., 2014). These brain regions engaged during auditory selective spatial attention are also engaged to some degree during non-spatial attention. However, there are also differences in activity, depending on exactly what acoustic feature guides attentional focus (e.g., see Hill & Miller, 2010; Lee et al., 2013). Still, there is no definitive list of which sound attributes or statistics can be used to focus auditory attention, or exactly which cortical regions are engaged by attention to specific acoustic features.

Failures of Selective Attention

There are several contextual factors that interfere with selective attention. First, the presence of acoustically similar distractors interferes with a variety of tasks that depend on selective attention (e.g., see the discussion in Durlach et al., 2003; Kidd et al., 2008). Second, uncertainty about what sound is a target and what is a distractor can lead to suboptimal selective attention (e.g., because the target or distractor is random and unpredictable, it can be difficult to filter out the distractor). Failures of attention due to these factors ("similarity" and "uncertainty") have been intensively studied and are discussed as different forms of IM (Durlach et al., 2003; Kidd et al., 2008).

For example, detection and discrimination of a target tone is much more difficult in the presence of simultaneous tones that are remote in frequency (Neff & Green, 1987) and/or tones that may overlap with the target in frequency but not in time (Watson et al., 1975, 1976), even though such interfering tones do not alter the target's representation at the level of the auditory nerve. In addition, performance suffers if uncertainty is increased by roving the characteristics of the target or distractors from trial to trial (e.g., Green, 1961; Spiegel et al., 1981; Neff & Dethlefs, 1995; see the review in Kidd et al., 2008). In the case of spatial attention, it seems intuitive that attending to one location might fail when competing sounds are added at nearby locations (akin to visual "crowding"), but few past experiments have addressed this issue directly. Spatial uncertainty (i.e., not knowing where a target stream will come from) disrupts selective attention; for example, the ability to focus on and analyze a stream of speech is poorer when its spatial location changes compared to when its location is fixed and known (Kidd et al., 2005; Brungart & Simpson, 2007; Best et al., 2008). In addition, listeners make fewer errors when attending to a fixed and known voice than when the target voice characteristics change regularly. In such cases, performance can be improved by providing cues to prime the listener to focus on target features such as frequency,

spatial location, or voice (Darwin et al., 2003; Richards & Neff, 2004; Kidd et al., 2005; Brungart & Simpson, 2007).

The failures of attention described above often show up in performance measures as target–masker confusions or substitutions, indicating that the listener extracted meaning from the sound mixture he or she heard but analyzed the wrong auditory object (e.g., Brungart, 2001; Brungart et al., 2001; Ihlefeld & Shinn-Cunningham, 2008b). However, there is another stage at which failures of attention can occur. As mentioned above, attention relies on the appropriate segregation of acoustic mixtures into well-formed objects. Perceptual segregation requires fine spectrotemporal details; if the sensory representation of sound is too muddled to support sound segregation, then selective attention can fail even when listeners know what to attend to (Shinn-Cunningham, 2008). As an example, listeners with hearing loss, especially those hearing through cochlear implants, receive highly distorted auditory inputs and are poor at segregating acoustic mixtures; consistent with this, the main complaint of such listeners is that they have trouble understanding sound in noisy backgrounds, where selective attention is necessary (see Shinn-Cunningham & Best, 2008, for a more detailed discussion of these ideas). Even listeners with good peripheral hearing may hear a sound mixture whose content is too chaotic to support source segregation, for instance, if there are too many sound sources in the scene or if there is significant reverberant energy distorting the spectrotemporal structure important for segregating sound sources (e.g., see Lee & Shinn-Cunningham, 2008; Mandel et al., 2010).

Dividing and Shifting Attention

While the basic filter theory of attention is built around the premise that attention operates to select one source and exclude others, there have been several attempts over the years to investigate the extent to which this selectivity is obligatory. In other words, to what extent can attention be divided between two or more auditory objects if that is the goal?

For relatively simple detection tasks, it seems that listeners can divide attention between two sound streams, with performance comparable to that achieved when monitoring a single stream. For example, when listeners monitor either two frequencies or two ears for target stimuli, detection in one channel seems to suffer only when a target occurs simultaneously in the other channel (Pashler, 1998). It is worth noting, however, that the ability to monitor multiple streams successfully may require more "effort" (a notoriously difficult thing to measure).

In the extensive literature on competing speech mixtures, there is little evidence that listeners actually divide attention. When listeners follow one talker, they appear to recall little about unattended talkers (Cherry, 1953). It is true that when listeners are instructed in advance to report back both of two competing messages, listeners can perform relatively well (Best et al., 2006; Gallun et al., 2007; Ihlefeld & Shinn-Cunningham, 2008a). However, it is not clear that this good performance indicates a

true sharing of attention across streams. One possibility is that attention can be divided to a point, when the stimuli are brief, when the two tasks are not demanding, and/or when the two tasks do not compete for a limited pool of processing resources (Gallun et al., 2007). Another possibility is that listeners process simultaneous inputs to the auditory system serially (Broadbent, 1954, 1956). When two sequences of digits are presented simultaneously to the two ears (or in two voices), listeners can recall all digits, but they first report the digits presented to one ear (or spoken by one voice) and then recall the content of the other stream (Broadbent, 1954, 1956). Broadbent postulated that simultaneous sensory inputs are stored temporarily via immediate auditory memory and then processed serially by a limited-capacity mechanism (Broadbent, 1957; see chapter 9 in Broadbent, 1958, and Lachter et al., 2004). A consequence of such a scheme is that the secondary message in the pair must be held in a volatile memory store while the primary message is processed.

Another question that has interested researchers is how easily and rapidly selective attention can be switched from one object to another when the focus of interest changes. There are many examples showing that there is a cost associated with switching auditory attention. Early experiments demonstrated deficits in recall of speech items when presented alternately to the two ears (Broadbent, 1954, 1956; Cherry & Taylor, 1954; Treisman, 1971). This cost is also apparent in more complex scenarios where listeners must switch attention on cue between multiple simultaneous streams of speech (Best et al., 2008). Costs of switching attention have also been demonstrated when the switch is from one voice to another (Larson & Lee, 2013; Lawo & Koch, 2014). The cost of switching attention is associated with the time required to disengage and reengage attention but may also come from an improvement in performance over time when listeners are able to hone the attentional filter more finely when they maintain focus on a single stream (e.g., see Best et al., 2008; Bressler et al., 2014).

Selective Attention in Complex Listening Scenarios

Over the last decade or so, there has been a surge of interest in studying attention in more natural, complex listening scenarios to try to strengthen the link between laboratory results and real-world behavior. Earlier improvements in technology facilitated this new research by providing sophisticated new ways to create complex auditory scenes like those encountered in everyday settings (e.g., see Carlile, 1996).

In most controlled experiments, listeners are asked to repeat back, verbatim, the contents of a target sentence or digit sequence. In contrast, in most verbal exchanges outside the lab, the exact wording of a message is irrelevant; only the meaning of the words is critical (albeit with some notable exceptions, such as understanding telephone numbers, etc.). Moreover, in typical conversations, that meaning must be tracked continuously as the conversation goes on. In the lab, however, messages are brief and often organized into trials where the subject has ample time to report back the content in between trials. Some recent experiments have attempted to bridge this gap, asking

listeners to maintain attention on speech that flows rapidly and continuously (Hafter, Xia, & Kalluri, 2013; Hafter, Xia, Kalluri, Poggesi, et al., 2013). Listeners were presented with competing stories and asked interpretive questions (i.e., requiring semantic, not just phonetic processing); the questions were most often about one of the stories (the target) but occasionally were about the competing story. Results revealed different kinds of limits on the processing of simultaneous speech from multiple talkers. For example, listeners only had the capacity to partially process one (but not two) competitors, and spatial attention limited that capacity to nearby competitors. Studies like this, which better recreate the pressures associated with continuously extracting meaning from ongoing conversations, are likely to produce new insights into how attention influences everyday communication.

In an attempt to understand the role of selective and divided attention in busy, natural listening scenarios, Eramudugolla and colleagues (2005) designed a novel "change deafness" paradigm. In a scene consisting of four to eight spatially separated natural sounds, they showed that when selective attention was directed in advance to one object, listeners were remarkably good at monitoring that object and detecting its disappearance in a subsequent exposure to the scene. However, in the absence of directed attention (i.e., when relying on divided attention) listeners were unable to reliably detect the disappearance of one of the objects.

Conversely, when listeners do focus attention selectively within a complex scene, it can leave them completely unaware of unusual or unexpected auditory events. For instance, one demonstration of "inattentional deafness" used an auditory analogue of the famous visual gorilla experiment. In the auditory version, listeners sustained their attention on one of two conversations in a simulated cocktail party. Under these conditions, many of the listeners failed to notice the presence of a man walking around the simulated scene repeatedly saying "I am a gorilla," despite both the prolonged duration of his message (19 seconds) and the fact that the message was audible and intelligible when attended (Dalton & Fraenkel, 2012). Inattentional deafness has also been demonstrated in complex musical scenes for nonmusicians and musicians alike (Koreimann et al., 2014). That is, focusing on one object can leave people unaware of the content of a competing object. In addition, focusing on one aspect of one object can leave people unaware of changes in other aspects of that object. When listeners are selectively focused on the content of an important message, they are often not even aware of a switch of the talker identity midway through the message (Vitevitch, 2003); moreover, the likelihood of listeners' noting such talker changes decreases as the demands of processing the lexical content of the message increases (Vitevitch & Donoso, 2011). While the first form of inattentional deafness (missing the presence of an unattended object) probably reflects the filtering out of unwanted information, so that its content is never processed, the second (missing discontinuities in one feature of a stream of information that is being processed) may be due to a failure to extract

and store in memory high-order aspects of an attended object (such as talker identity) when they are not critical to performance and other task demands are high. Further work is needed to explore how auditory memory interacts with perception in complex, demanding scenes (see Snyder & Gregg, 2011, for a recent review of such issues).

Integration

The expansive literature on auditory selective attention has come about because of an intuitive recognition of its importance in how we function in the world. Competing sounds are a feature of most everyday environments, and selective attention is critical for enabling us to navigate and communicate in common situations. Yet, there are some major distinctions between what we typically test in laboratory experiments and how listeners normally operate in the real world. One major difference is the kind of information listeners typically extract from auditory scenes. Almost no laboratory experiments test listeners' awareness of the ambience of a setting (busy indoor café or a forest in the height of spring?), yet such information is something we are constantly monitoring, even when we are not consciously aware that we are doing so. Similarly, as noted in the previous section, in the laboratory we often test how well listeners can report back the exact content of brief spoken messages, even though it is really the meaning of ongoing conversations that must be conveyed in the boardroom, the football field, the coffee shop, and so on. Of course, it is relatively easy to come up with objective scores for how well listeners can report back an exact sequence of words compared to testing how well they understood the meaning of a long verbal exchange. Yet to truly understand how selective attention affects our everyday interactions, we must quantify performance in more natural scenarios.

While there has been a lot of experimental work on failures of attention related to selection (e.g., studies of IM), there is surprisingly little known about how often these kinds of failures occur in the real world. In many laboratory paradigms, stimuli are manipulated to create scenes in which competing streams are unnaturally similar or unnaturally correlated in their timing. Yet naturally occurring sounds in the real world come from independent physical sources, so are generally distinct, temporally uncorrelated, from different locations, and unrelated in all other dimensions. Moreover, other nonacoustic information, such as visual cues, reinforces these acoustic cues and reduces uncertainty about where and when to attend to extract a particular source. These (and other) properties of real-world listening scenarios reduce the likelihood of confusion compared to what we researchers often create in a "good" experimental design. Despite this, introspectively, one can think of occasions where a competing conversation "intruded" into the conversation of interest. Imagine trying to listen to the song of one bird in the dawn chorus (where the similarity of the individual bird songs makes it difficult to segregate any one physical source from the rest). Moreover, there is evidence that certain populations, such as children and the elderly, have particular difficulties

in suppressing unwanted sounds (e.g., see Tun et al., 2002; Elliott, 2002). A recent study showed that performance on a laboratory test that measures people's ability to maintain attention on an auditory task in the presence of a distractor can be predicted by performance on a questionnaire that rates everyday distractibility (the Cognitive Failures Questionnaire; Murphy & Dalton, 2014). These studies hint that even though laboratory experiments rarely reflect the kinds of challenges facing ordinary listeners in ordinary settings, the factors studied in the laboratory nonetheless affect our ability to communicate in everyday life.

We touched above on object formation and its role in selective attention. In the real world it is very likely that failures of object formation are a significant contributor to failures of selective attention. In real-world listening environments, the defining features of sound objects are often degraded by interactions with other sounds and reflections from walls and other surfaces. Moreover, it is common for sounds with similar spectra to occur simultaneously, causing masking in the peripheral representation of sound (EM) to affect everyday perception. These factors are exacerbated in listeners with hearing loss and in cochlear implant users because of their reduced peripheral resolution (Shinn-Cunningham & Best, 2008).

The challenges associated with dividing and switching attention discussed above are crucially important for understanding real-world behavior. To return to our starting point of the cocktail party problem, it is clear that most social interactions involve not just paying attention to one talker in the presence of unwanted distractors but rather dynamically redirecting attention to different talkers with different vocal characteristics, who are generally at different locations or even moving around, all in unpredictable and unexpected ways. Indeed, the more boisterous, and "social" a conversation is, the more likely it will involve unexpected interruptions by some animated, engaged participant responding to a previous talker's point. Thus, attention must constantly be divided and shifted in order to follow the flow of conversation and participate meaningfully in social settings, stressing our perceptual skills in ways rarely tested in the laboratory.

A final issue worth considering is the possible interplay between technology and auditory selective attention. Recent advances in hearing devices provide a number of new opportunities for aiding listeners who have difficulty communicating in normal social scenes (e.g., see Edwards, 2007; Kidd et al., 2013). Our scientific understanding of auditory attention should guide the development of such new devices and focus resources on those aspects of the problem that lead to failures of communication—and to social isolation. On the other side of the spectrum, new technologies can also bring new challenges related to auditory selective attention that we are only just beginning to understand. For example, it is clear that focused attention to portable music players and cell phones has consequences for awareness of other critical sounds in the environment (sirens, alarms, approaching cars) that can interfere with fundamental tasks such as walking and driving.

Summary and Future Directions

Some of the earliest research on attention came out of work on auditory selective attention. While visual researchers embraced this early work, many psychoacousticians instead focused on bottom-up processing limitations of the auditory system. Over the last 15 to 20 years, auditory researchers have turned back to studying the importance of selective attention in auditory perception, basing much of their work on the breakthroughs from visual science.

Many aspects of auditory selective attention seem to operate analogously to how selective attention operates in vision. For instance, there is a close relationship between object formation and attention in both modalities. In line with the idea that the "unit" of selective attention is an object, listeners show little evidence for being able to truly divide attention; instead, they appear to switch attention rapidly between sources and to use memory to fill in the gaps. Many other perceptual "failures" in complex listening scenarios are also consistent with auditory attention's analyzing one, and only one, object at a time. For instance, the idea that listeners are unable to monitor multiple sources simultaneously in a complex scene explains both change deafness and inattentional deafness.

Various auditory features both support auditory scene analysis (source segregation) and serve as perceptual dimensions that can be used to direct attention, selecting out a target object from a complex acoustic scene. These dimensions include frequency content, location, timbre, pitch, and rhythm; however, there is no well-defined list of discrete features that can be used to direct attention. This is one area where much work remains.

While many of the basic principles of selective attention are similar in auditory and visual perception, one striking difference is in the importance of time in auditory perception. Auditory sources are inherently temporal; auditory information can be extracted only by listening to a sound through time. In line with this, an auditory object can be resolved from an auditory scene only by analyzing the spectrotemporal content of sound across a range of timescales, from milliseconds to tens of seconds. Because of this, and because acoustically, sound combines additively before entering the ears, auditory selective attention is often limited by difficulties with *segregating* an object of interest from a scene. Source segregation is even harder in the presence of reverberant energy and background noise; any degradation of the sensory representation of sound in the auditory periphery, such as from hearing loss, exacerbates the problems of source segregation. Thus, in everyday settings, selective auditory attention often fails because of failures of object formation. Given the important role that time plays in auditory information, the dynamics of attention (e.g., the time course for attention to be focused, reoriented, maintained, etc.) is, if anything, more critical than in other sensory dimensions. Yet, relatively few studies have tackled the problem of exploring the dynamics of selective auditory attention.

The demands we face every day are very different than those tested in most laboratories. In the real world, auditory scenes are unpredictable and ongoing. Real-world social settings require listeners to keep up and constantly extract meaning from ongoing conversations that are full of unpredictable interruptions and shifts and that often take place in noisy, reverberant settings. In addition, competition from nonauditory modalities is inevitable in busy everyday environments (think about how distracting a TV screen can be in a pub or a restaurant when you are trying to listen to a conversation). New insights will come from expanding our research to better match the complexities of everyday settings. Insights from such studies will be especially helpful in understanding the difficulties faced by various special populations, from listeners with hearing impairment to children to veterans with traumatic brain injury to listeners with deficits in cognitive function.

Box 5.1
Key Points

- Selective auditory attention cannot operate unless listeners can segregate the acoustic mixture reaching the ears into constituent sound sources.
- Listeners can focus selective auditory attention on a sound object based on any number of features, from location to timbre to talker characteristics.
- Because sound conveys information through time, it is critical to study the dynamics of auditory attention.
- While laboratory experiments have demonstrated many of the key factors important for selective auditory attention, real-world settings require listeners to maintain awareness in dynamic and complicated sound mixtures and to keep up with ongoing, unpredictable conversations, unlike typical controlled experimental conditions.

Box 5.2
Outstanding Issues

- Further research is needed to explore and identify the discrete auditory features that can be used to direct attention.
- What is the time course for attention in the auditory domain to be focused, reoriented, maintained?
- How do the principles of auditory attention identified in the laboratory scale up to the more complex environments characteristic of our everyday lives?

References

Alain, C., Arnott, S. R., & Picton, T. W. (2001). Bottom-up and top-down influences on auditory scene analysis: Evidence from event-related brain potentials. *Journal of Experimental Psychology: Human Perception and Performance, 27*, 1072–1089.

Alain, C., & Woods, D. L. (1997). Attention modulates auditory pattern memory as indexed by event-related brain potentials. *Psychophysiology, 34*, 534–546.

Arbogast, T. L., & Kidd, G., Jr. (2000). Evidence for spatial tuning in informational masking using the probe-signal method. *Journal of the Acoustical Society of America, 108*, 1803–1810.

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., Ozmeral, E. J., Kopčo, N., & Shinn-Cunningham, B. G. (2008). Object continuity enhances selective auditory attention. *Proceedings of the National Academy of Sciences of the United States of America, 105*, 13174–13178.

Bregman, A. S. (1990). *Auditory scene analysis: The perceptual organization of sound*. Cambridge, MA: MIT Press.

Bressler, S., Masud, S., Bharadwaj, H., & Shinn-Cunningham, B. (2014). Bottom-up influences of voice continuity in focusing selective auditory attention. *Psychological Research, 78*, 349–360.

Broadbent, D. E. (1954). The role of auditory localization in attention and memory span. *Journal of Experimental Psychology, 47*, 191–196.

Broadbent, D. E. (1956). Successive responses to simultaneous stimuli. *Quarterly Journal of Experimental Psychology, 8*, 145–152.

Broadbent, D. E. (1957). Immediate memory and simultaneous stimuli. *Quarterly Journal of Experimental Psychology, 9*, 1–11.

Broadbent, D. E. (1958). *Perception and communication*. New York: Pergamon Press.

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

Brungart, D. S., & Simpson, B. D. (2007). Cocktail party listening in a dynamic multitalker environment. *Perception & Psychophysics, 69*, 79–91.

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.

Carlile, S. (1996). *Virtual auditory space: Generation and applications*. New York: R.G. Landes.

Carlyon, R. P. (2004). How the brain separates sounds. *Trends in Cognitive Sciences, 8*, 465–471.

Carlyon, R. P., Plack, C. J., Fantini, D. A., & Cusack, R. (2003). Cross-modal and non-sensory influences on auditory streaming. *Perception, 32*, 1393–1402.

Chait, M., de Cheveigne, A., Poeppel, D., & Simon, J. Z. (2010). Neural dynamics of attending and ignoring in human auditory cortex. *Neuropsychologia, 48*, 3262–3271.

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

Cherry, E. C., & Taylor, W. K. (1954). Some further experiments upon the recognition of speech, with one and with two ears. *Journal of the Acoustical Society of America, 26*, 554–559.

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

Corteen, R. S., & Wood, B. (1972). Autonomic responses to shock-associated words in an unattended channel. *Journal of Experimental Psychology, 94*, 308–313.

Culling, J. F., Hodder, K. I., & Toh, C. Y. (2003). Effects of reverberation on perceptual segregation of competing voices. *Journal of the Acoustical Society of America, 114*, 2871–2876.

Cusack, R., & Carlyon, R. P. (2003). Perceptual asymmetries in audition. *Journal of Experimental Psychology: Human Perception and Performance, 29*, 713–725.

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.

Dalton, P., & Fraenkel, N. (2012). Gorillas we have missed: Sustained inattentional deafness for dynamic events. *Cognition, 124*, 367–372.

Darwin, C. J., Brungart, D. S., & Simpson, B. D. (2003). Effects of fundamental frequency and vocal-tract length changes on attention to one of two simultaneous talkers. *Journal of the Acoustical Society of America, 114*, 2913–2922.

Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual Review of Neuroscience, 18*, 193–222.

Devergie, A., Grimault, N., Tillmann, B., & Berthommier, F. (2010). Effect of rhythmic attention on the segregation of interleaved melodies. *Journal of the Acoustical Society of America, 128*, EL1–EL7.

Durlach, N. I., Mason, C. R., Shinn-Cunningham, B. G., Arbogast, T. L., Colburn, H. S., & Kidd, G., Jr. (2003). Informational masking: Counteracting the effects of stimulus uncertainty by decreasing target–masker similarity. *Journal of the Acoustical Society of America, 114*, 368–379.

Edwards, B. (2007). The future of hearing aid technology. *Trends in Amplification, 11*, 31–45.

Elhilali, M., Ma, L., Micheyl, C., Oxenham, A. J., & Shamma, S. A. (2009). Temporal coherence in the perceptual organization and cortical representation of auditory scenes. *Neuron, 61*, 317–329.

Elliott, E. M. (2002). The irrelevant-speech effect and children: Theoretical implications of developmental change. *Memory & Cognition, 30*, 478–487.

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

Freyman, R. L., Helfer, K. S., McCall, D. D., & Clifton, R. K. (1999). The role of perceived spatial separation in the unmasking of speech. *Journal of the Acoustical Society of America, 106*, 3578–3588.

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.

Gallun, F. J., Mason, C. R., & Kidd, G., Jr. (2007). Task-dependent costs in processing two simultaneous auditory stimuli. *Perception & Psychophysics, 69*, 757–771.

Gray, G., & Wedderburn, A. (1960). Grouping strategies with simultaneous stimuli. *Quarterly Journal of Experimental Psychology, 12*, 180–185.

Green, D. M. (1961). Detection of auditory sinusoids of uncertain frequency. *Journal of the Acoustical Society of America, 33*, 897–903.

Greenberg, G. Z., & Larkin, W. D. (1968). Frequency-response characteristic of auditory observers detecting signals of a single frequency in noise: The probe-signal method. *Journal of the Acoustical Society of America, 44*, 1513–1523.

Griffiths, T. D., & Warren, J. D. (2004). What is an auditory object? *Nature Reviews. Neuroscience, 5*, 887–892.

Hafter, E. R., Xia, J., & Kalluri, S. (2013). A naturalistic approach to the cocktail party problem. *Advances in Experimental Medicine and Biology, 787*, 527–534.

Hafter, E. R., Xia, J., Kalluri, S., Poggesi, R., Hansen, C., & Whiteford, K. (2013). Attentional switching when listeners respond to semantic meaning expressed by multiple talkers. *Journal of the Acoustical Society of America, 133*, 3381.

Hansen, J. C., & Hillyard, S. A. (1980). Endogenous brain potentials associated with selective auditory attention. *Electroencephalography and Clinical Neurophysiology, 49*, 277–290.

Hill, K. T., & Miller, L. M. (2010). Auditory attentional control and selection during cocktail party listening. *Cerebral Cortex, 20*, 583–590.

Hupe, J. M., Joffo, L. M., & Pressnitzer, D. (2008). Bistability for audiovisual stimuli: Perceptual decision is modality specific. *Journal of Vision, 8*(7), 1, 1–15.

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.

Jones, M. R. (1976). Time, our lost dimension: Toward a new theory of perception, attention, and memory. *Psychological Review, 83,* 323–355.

Jones, M. R., Kidd, G., & Wetzel, R. (1981). Evidence for rhythmic attention. *Journal of Experimental Psychology: Human Perception and Performance, 7,* 1059–1073.

Kastner, S., & Ungerleider, L. G. (2001). The neural basis of biased competition in human visual cortex. *Neuropsychologia, 39,* 1263–1276.

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., Favrot, S., Desloge, J. G., Streeter, T. M., & Mason, C. R. (2013). Design and preliminary testing of a visually guided hearing aid. *Journal of the Acoustical Society of America, 133,* EL202–EL207.

Kidd, G., Jr., Mason, C. R., Deliwala, P. S., Woods, W. S., & Colburn, H. S. (1994). Reducing informational masking by sound segregation. *Journal of the Acoustical Society of America, 95,* 3475–3480.

Kidd, G., Jr., Mason, C. R., Richards, V. M., Gallun, F. J., & Durlach, N. I. (2008). Informational masking. In W. Yost, A. N. Popper, & R. R. Fay (Eds.), *Springer handbook of auditory research: Vol. 29. Auditory perception of sound sources* (pp. 143–189). New York: Springer.

Kitterick, P. T., Clarke, E., O'Shea, C., Seymour, J., & Summerfield, A. Q. (2013). Target identification using relative level in multi-talker listening. *Journal of the Acoustical Society of America, 133,* 2899–2909.

Koreimann, S., Gula, B., & Vitouch, O. (2014). Inattentional deafness in music. *Psychological Research, 78,* 304–312.

Lachter, J., Forster, K. I., & Ruthruff, E. (2004). Forty-five years after Broadbent (1958): Still no identification without attention. *Psychological Review, 111,* 880–913.

Larson, E., & Lee, A. K. C. (2013). Influence of preparation time and pitch separation in switching of auditory attention between streams. *Journal of the Acoustical Society of America, 134,* EL165–EL171.

Lawo, V., & Koch, I. (2014). Dissociable effects of auditory attention switching and stimulus-response compatibility. *Psychological Research, 78,* 379–386.

Lee, A. K. C., Larson, E., Maddox, R. K., & Shinn-Cunningham, B. G. (2014). Using neuroimaging to understand the cortical mechanisms of auditory selective attention. *Hearing Research, 307,* 111–120.

Lee, A. K. C., Rajaram, S., Xia, J., Bharadwaj, H., Larson, E., Hamalainen, M., et al. (2013). Auditory selective attention reveals preparatory activity in different cortical regions for selection based on source location and source pitch. *Frontiers in Neuroscience, 6*, 190.

Lee, A. K. C., & Shinn-Cunningham, B. G. (2008). Effects of reverberant spatial cues on attention-dependent object formation. *Journal of the Association for Research in Otolaryngology, 9*, 150–160.

Mackay, D. G. (1973). Aspects of the theory of comprehension, memory and attention. *Quarterly Journal of Experimental Psychology, 25*, 22–40.

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.

Maddox, R. K., Billimoria, C. P., Perrone, B. P., Shinn-Cunningham, B. G., & Sen, K. (2012). Competing sound sources reveal spatial effects in cortical processing. *PLoS Biology, 10*(5), e1001319.

Maddox, R. K., & Shinn-Cunningham, B. G. (2012). Influence of task-relevant and task-irrelevant feature continuity on selective auditory attention. *Journal of the Association for Research in Otolaryngology, 13*, 119–129.

Mandel, M. I., Bressler, S., Shinn-Cunningham, B., & Ellis, D. P. W. (2010). Evaluating source separation algorithms with reverberant speech. *IEEE Transactions on Audio Speech and Language Processing, 18*, 1872–1883.

Mesgarani, N., & Chang, E. F. (2012). Selective cortical representation of attended speaker in multi-talker speech perception. *Nature, 485*, 233–236.

Middlebrooks, J. C., & Bremen, P. (2013). Spatial stream segregation by auditory cortical neurons. *Journal of Neuroscience, 33*, 10986–11001.

Mondor, T. A., & Zatorre, R. J. (1995). Shifting and focusing auditory spatial attention. *Journal of Experimental Psychology: Human Perception and Performance, 21*, 387–409.

Moray, N. (1959). Attention in dichotic listening: Affective cues and the influence of instructions. *Quarterly Journal of Experimental Psychology, 11*, 56–60.

Murphy, S., & Dalton, P. (2014). Ear-catching? Real-world distractibility scores predict susceptibility to auditory attentional capture. *Psychonomic Bulletin & Review, 21*, 1209–1213.

Neff, D. L., & Dethlefs, T. M. (1995). Individual differences in simultaneous masking with random-frequency, multicomponent maskers. *Journal of the Acoustical Society of America, 98*, 125–134.

Neff, D. L., & Green, D. M. (1987). Masking produced by spectral uncertainty with multicomponent maskers. *Perception & Psychophysics, 41*, 409–415.

Oh, E. L., & Lutfi, R. A. (1998). Nonmonotonicity of informational masking. *Journal of the Acoustical Society of America, 104*, 3489–3499.

Pashler, H. E. (1998). *The psychology of attention*. Cambridge, MA: MIT Press.

Picton, T. W., & Hillyard, S. A. (1974). Human auditory evoked potentials: II. Effects of attention. *Electroencephalography and Clinical Neurophysiology, 36*, 191–199.

Quinlan, P. T., & Bailey, P. J. (1995). An examination of attentional control in the auditory modality: Further evidence for auditory orienting. *Perception & Psychophysics, 57*, 614–628.

Richards, V. M., & Neff, D. L. (2004). Cuing effects for informational masking. *Journal of the Acoustical Society of America, 115*, 289–300.

Scharf, B., Quigley, S., Aoki, C., Peachey, N., & Reeves, A. (1987). Focused auditory attention and frequency selectivity. *Perception & Psychophysics, 42*, 215–223.

Shamma, S. A., Elhilali, M., & Micheyl, C. (2011). Temporal coherence and attention in auditory scene analysis. *Trends in Neurosciences, 34*, 114–123.

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

Shinn-Cunningham, B. G., & Best, V. (2008). Selective attention in normal and impaired hearing. *Trends in Amplification, 12*, 283–299.

Snyder, J. S., & Gregg, M. K. (2011). Memory for sound, with an ear toward hearing in complex auditory scenes. *Attention, Perception & Psychophysics, 73*, 1993–2007.

Snyder, J. S., Gregg, M. K., Weintraub, D. M., & Alain, C. (2012). Attention, awareness, and the perception of auditory scenes. *Frontiers in Psychology, 3*, 15.

Spence, C., Nicholls, M. E. R., & Driver, J. (2001). The cost of expecting events in the wrong sensory modality. *Perception & Psychophysics, 63*, 330–336.

Spence, C. J., & Driver, J. (1994). Covert spatial orienting in audition: Exogenous and endogenous mechanisms. *Journal of Experimental Psychology: Human Perception and Performance, 20*, 555–574.

Spiegel, M. F., Picardi, M. C., & Green, D. M. (1981). Signal and masker uncertainty in intensity discrimination. *Journal of the Acoustical Society of America, 70*, 1015–1019.

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

Tark, K. J., & Curtis, C. E. (2013). Deciding where to look based on visual, auditory, and semantic information. *Brain Research, 1525*, 26–38.

Teki, S., Chait, M., Kumar, S., Shamma, S., & Griffiths, T. D. (2013). Segregation of complex acoustic scenes based on temporal coherence. *eLife, 2*, e00699.

Treisman, A. M. (1960). Contextual cues in selective listening. *Quarterly Journal of Experimental Psychology, 12*, 157–167.

Treisman, A. M. (1971). Shifting attention between the ears. *Quarterly Journal of Experimental Psychology, 23*, 157–167.

Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology, 12*, 97–136.

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.

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

Vitevitch, M. S., & Donoso, A. (2011). Processing of indexical information requires time: Evidence from change deafness. *Quarterly Journal of Experimental Psychology, 64*, 1484–1493.

Watson, C. S. (2005). Some comments on informational masking. *Acta Acustica united with Acustica, 91*, 502–512.

Watson, C. S., Kelly, W. J., & Wroton, H. W. (1976). Factors in the discrimination of tonal patterns. II. Selective attention and learning under various levels of stimulus uncertainty. *Journal of the Acoustical Society of America, 60*, 1176–1186.

Watson, C. S., Wroton, H. W., Kelly, W. J., & Benbassat, C. A. (1975). Factors in the discrimination of tonal patterns. I. Component frequency, temporal position, and silent intervals. *Journal of the Acoustical Society of America, 57*, 1175–1185.

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.

Wright, B. A., & Fitzgerald, M. B. (2004). The time course of attention in a simple auditory detection task. *Perception & Psychophysics, 66*, 508–516.