Memory for sound, with an ear toward hearing in complex auditory scenes

Joel S. Snyder · Melissa K. Gregg

Published online: 2 August 2011
© Psychonomic Society, Inc. 2011

Abstract An area of research that has experienced recent growth is the study of memory during perception of simple and complex auditory scenes. These studies have provided important information about how well auditory objects are encoded in memory and how well listeners can notice changes in auditory scenes. These are significant developments because they present an opportunity to better understand how we hear in realistic situations, how higher-level aspects of hearing such as semantics and prior exposure affect perception, and the similarities and differences between auditory perception and perception in other modalities, such as vision and touch. The research also poses exciting challenges for behavioral and neural models of how auditory perception and memory work.

Keywords auditory scene analysis · auditory memory · change deafness · attention

One of the major goals of research on auditory cognition is explaining how we perceive and remember realistic situations, such as a cocktail party in which competing sound sources are present (Cherry, 1953). Such complex situations are typical of everyday life, but they also pose special difficulty for populations with hearing problems, such as normal aging adults and those with hearing loss (Alain, Dyson & Snyder, 2006; Committee on Hearing, 1988; Schneider, Daneman & Pichora-Fuller, 2002). The process of hearing in situations such as a cocktail party is called *auditory scene analysis* (Bregman, 1990), analogous to the study of *visual scene analysis*, which explains how

the visual system is able to parse a complex scene into its constituent parts or objects (see Chap. 4 of Wolfe et al., 2008). A large body of research on auditory scene analysis has provided many insights into how we perceptually organize mixtures of multiple sound sources into objects or streams, although little research has studied perception of recognizable objects that have semantic meaning (Alain, 2007; Bregman, 1990; Carlyon, 2004; Darwin, 1997; Micheyl et al., 2007; Moore & Gockel, 2002; Snyder & Alain, 2007).

The definition of an auditory object (or *stream*) has not always been clear in the literature, despite the frequency of the use of this construct and its obvious importance for conceptualizing auditory perception and memory in real-world situations. According to Bregman (1990), an auditory stream is the “natural” object of attention because “perception should be concerned, then, with properties calculated on the basis of an appropriate grouping of information, rather than with properties calculated from isolated flashes of light or squeaks of a sound.” Similarly, Alain and Arnott (2000) posited that auditory objects are components of a scene that have been analyzed according to perceptual grouping principles, and Kubovy and Van Valkenburg (2001) assumed that an auditory object is something that has been perceptually segregated as figure from ground (see also Griffiths & Warren, 2004). Related definitions have been offered for visual objects as entities that emerge when no more Gestalt grouping can be performed (see Feldman, 2003). All of these definitions are similar in that they imply that auditory attention is concerned with global sound configurations, rather than with individual acoustic features (Dyson & Ishfaq, 2008; Melara & Marks, 1990; Mondor, Zatorre & Terrio, 1998). For our purposes, we will define an auditory *object* or *stream* as equivalently meaning a single perceptual entity that has resulted from grouping and attention processes. Therefore, we define a *scene* as a

J. S. Snyder (✉) · M. K. Gregg
Department of Psychology, University of Nevada Las Vegas,
4505 South Maryland Parkway, Box 455030, Las Vegas, NV,
89154-5030, USA
e-mail: Joel.Snyder@unlv.edu

situation in which more than one such object or stream is perceived. Note that these definitions are inherently listener centered and do not necessarily agree with the objective state of affairs in the physical world. However, it is also important to note that when experiments present recognizable sounds to listeners (e.g., a dog barking and the sound of an airplane), it is generally assumed that the listener will group the appropriate sound components in a scene with other components coming from the same source.

According to Bregman (1990), we accomplish auditory scene analysis by using fairly automatic *primitive* organizational mechanisms and more knowledge-driven or attention-driven *schema-based* mechanisms. Schema-based mechanisms—the use of attention, prior memories, and knowledge to organize auditory scenes—are much less understood. In fact, few studies have focused on understanding auditory memory for scenes at all. This is in spite of the fact that memory is likely to play an important role in the experiencing of everyday auditory scenes. For example, being able to recognize whether one is encountering a familiar person or object might rely to a significant degree on auditory information encoded in long-term memory, in addition to cues in other modalities, such as vision. Also, being able to notice that a sound source has arrived or departed may be a valuable ability, especially for parts of the scene that are not currently in the field of view. Finally, there are likely to be many real-world examples in which implicit auditory memory plays an important role in sensory and perceptual encoding processes, especially for sounds that are extended in time (e.g., a melody or a sentence) and for sounds that repeat within a short period of time (e.g., a bird calling or a car honking).

Recent studies have helped define some of the important issues in the field of auditory memory that are likely to guide further research in this area. An important issue to understand from a basic science perspective is what abilities and limitations are characteristic of various types of auditory memory and what neurocomputational mechanisms account for observed patterns of behavior. A secondary issue that has received a fair amount of attention is the extent to which auditory memory is similar to or different from memory in other sensory modalities. Comparisons with vision are especially helpful because of the numerous studies of visual memory and the existence of well-developed theories of visual processing. It should be mentioned that some differences across modalities are surely to be expected due to the different information encoded in the periphery and subcortical areas and the greater amount of subcortical processing occurring in the auditory domain (Nelken, 2004). However, at the cortical level, where many memory functions are likely to occur, it might be expected that rather similar computations are carried out in modality-specific sensory areas (e.g., DeWeese, Hromádka & Zador,

2005; Shamma, 2001). It is also possible that certain memory functions are implemented in shared brain circuits, regardless of the stimulated modality. A third issue of importance is the extent to which studying auditory memory for simple acoustic patterns can help us understand memory for more complex auditory objects and scenes. The present article will illuminate these issues by focusing on explicit memory, the ability to encode and recall information about experienced events and learned knowledge or facts (Squire, Stark & Clark 2004; Squire & Zola-Morgan, 1988; Thompson & Kim, 1996).

A comprehensive review of research on auditory memory is not possible here, which has led us to focus on recent research about explicit memory for recognizable auditory objects and scenes. We will begin by discussing recent work on detecting changes in auditory scenes—in particular, the phenomenon of *change deafness*. Although this is a new topic of inquiry, the research published to date allows us to address all of the major theoretical issues outlined above. The study of change deafness also nicely illustrates the interaction between auditory memory and many other aspects of cognition, including attention, awareness, and the encoding of auditory objects. Next, we will summarize research on *recognition* and *recall* of previously presented sounds. Where possible, we will discuss studies using brain measurement techniques, especially when such studies help identify the level of processing or type of mechanism involved in a particular type of auditory memory. We will not extensively survey studies providing evidence for distinct auditory stores and the basic mechanisms for remembering and forgetting, since reviews of this topic have already been published (Cowan, 1984; Demany & Semal, 2008; McLachlan & Wilson, 2010; Näätänen & Winkler, 1999; Nairne, 2003). We will also not focus on the various forms of implicit memory, such as perceptual learning (Dahmen & King, 2007; Jaaskelainen, Ahveninen, Belliveau, Raji & Sams, 2007; Samuel & Kraljic, 2009; Weinberger, 2004; Wright & Zhang, 2009). Finally, we will not focus on how memory for different materials might differ from each other (i.e., speech vs. music vs. other environmental sounds).

Change detection

A potentially informative line of research for understanding the abilities and limitations of memory for natural scenes is the recent work on change detection. Auditory and visual change detection research has shown that listeners are remarkably poor at detecting changes to natural scenes. For example, around two thirds of participants fail to notice a change in the identity of a conversation partner when the conversation is interrupted by a brief visual occlusion

(Simons & Levin, 1998). This phenomenon, called *change blindness* in the visual domain and *change deafness* in the auditory domain, is particularly relevant to the topic of memory, because the occurrence of both suggests that memory during scene perception is quite error prone. Change blindness and change deafness raise several intriguing questions regarding the nature of memory during natural scene perception: Does memory play only a minimal role in natural scene perception? Is the type of memory used during natural scene perception not equipped to maintain the details of a scene? Or does the interaction of memory with some other cognitive process hinder the ability to retain detail during natural scene perception?

The visual phenomenon of change blindness has been demonstrated with a range of stimuli from simple shapes (e.g., Luck & Vogel, 1997) to real-life social interactions (Simons & Levin, 1998). Two of the most common experimental paradigms in change blindness research are the one-shot technique and the flickering paradigm. In the one-shot technique, presentation of a scene is followed by an interruption and then either the same or a modified scene (e.g., Luck & Vogel, 1997). The observer's task is to report whether the two scenes are the same or different. In the flickering paradigm, two slightly different images are flickered back and forth until the participant can report the change (e.g., Rensink, O'Regan & Clark, 1997). The rate at which change blindness occurs varies across studies, although claims of change blindness typically rely on at least a 30% error on change trials—that is, saying “same” when there was in fact a change. Recent research indicates that individuals also miss substantial tactile (Auvray, Gallace, Hartcher-O'Brien, Tan & Spence, 2008; Gallace, Tan & Spence 2006) and auditory (e.g., Eramudugolla, Irvine, McAnally, Martin & Mattingley 2005) changes at rates comparable to or greater than those observed in the corresponding visual work.

The auditory phenomenon of change deafness has been demonstrated with speech (e.g., Vitevitch, 2003), environmental sounds (e.g., Eramudugolla et al., 2005), and musical stimuli (e.g., Agres & Krumhansl, 2008; Trainor & Trehub, 1992). Change deafness was actually demonstrated as early as the work of Cherry (1953), who showed that changes to an unattended stream of auditory input (such as a change of the speaker's identity) are often missed while shadowing a spoken message presented to an attended stream of auditory input (see also Sinnett, Costa & Soto-Faraco, 2006; Vitevitch, 2003). As is discussed in the following sections, recent studies using auditory paradigms that more closely resemble those in the change blindness work have definitively established the frequent occurrence of change deafness. Recent studies using the one-shot technique in audition (e.g., Eramudugolla et al., 2005; Gregg & Samuel, 2008, 2009) have shown that

listeners often miss changes to environmental objects, such as a dog barking changing to a piano tune. The small body of change deafness research that has been done indicates that listeners typically miss changes anywhere from about 30% (Pavani & Turatto, 2008) to 50% (Gregg & Samuel, 2008) of the time.

Comparing change deafness and change blindness

Similarities The use of similar paradigms across modalities makes this line of work particularly useful for an analysis of the similarities and differences of auditory and visual memory. One common finding across both visual and auditory research is that attention enhances change detection ability. Visual change detection improves when a verbal cue directs attention to the area of the picture that changes from flicker to flicker. For example, participants need about 17 alternations to see the change without a cue but only about 5 alternations to see the change with a cue (Rensink et al., 1997). Change deafness also is reduced with directed attention to the changing object. Eramudugolla et al. (2005) presented environmental sounds using the one-shot technique. They presented a 5-sec scene of sounds, followed by a burst of white noise, and then another 5-sec scene that was either the same or different. On *different* trials, either an object from scene 1 was deleted in scene 2, or two objects switched spatial locations from scene 1 to scene 2. The experimental task was to report whether the two scenes were the *same* or *different*. Eramudugolla et al. found that substantial change deafness occurred (nearly 50% for scene sizes of eight objects). However, when attention was directed to the object that could change via a verbal cue, change detection performance was nearly perfect (see also inattention studies; Sinnett et al., 2006; Vitevitch, 2003).

Given the large effect attention has on change detection, one logical theory of why change blindness/deafness occurs is that the changes are outside the focus of attention (e.g., Eramudugolla et al., 2005; Irwin & Andrews, 1996). Although the attention-based theories are likely to be partly correct about why change detection failures occur, one caveat is that change blindness has been demonstrated despite focused attention to the changing object. For example, Triesch, Ballard, Hayhoe and Sullivan (2003) found that participants rarely noticed changes to an object that was being attended to immediately before and after the change. In this study, participants were required to pick up and organize bricks in a virtual reality environment. Triesch et al. changed the size of a brick that participants were actively moving while participants made a saccade. Detection of the change was quite poor, although change detection performance improved significantly when the feature of the object that changed was task relevant (i.e., when participants were required to organize the bricks

according to size). Although a similar paradigm has not yet been applied to the auditory modality, the finding that attention is not always sufficient to detect visual changes suggests that other cognitive processes, such as memory during natural scene perception, may be especially important for detecting changes. This is, of course, an assumption based on the similar findings across modalities in change detection tasks that should be tested empirically.

Another common conclusion drawn from change blindness/deafness research is that failures to detect changes do not necessarily reflect a failure to encode objects in scenes. For example, Mitroff, Simons and Levin (2004) measured change detection and object encoding using the one-shot technique with simple visual objects. Their basic paradigm is presented in Fig. 1. As can be seen in the figure, they presented a 1-s visual scene, followed by a 350-ms blank screen, and then another 1-s scene that was either the same as or different from the first scene. Participants performed a change detection task, followed by an object-encoding task, by indicating which of two objects they had seen in scene 1 and/or scene 2. Object-encoding performance was better on trials on which the change was successfully detected, but accuracy on the object-encoding task was still significantly better than chance on trials on which change blindness occurred (the error rate on the object-encoding questions was only 26% in Experiment 1, despite a change detection error rate of about 34% on change trials).

Gregg and Samuel (2008) used a similar approach in the auditory domain and found that auditory objects also are well encoded, despite the fact that participants often miss changes in the set of objects (the error rate on the object-encoding task was only 28%, despite an error rate of 53% on the change detection task). Unlike in the study by Mitroff et al. (2004), however, Gregg and Samuel (2008) did not measure object encoding and change detection on

the same trials, so it was not possible to determine whether successful object encoding tended to co-occur with successful change detection. Gregg and Samuel (2008) also examined the effects of acoustic cues on change detection and object-encoding performance. They found that the acoustics of a scene were a critical determinant of change deafness; as can be seen in Fig. 2, performance improved when the object that changed was more acoustically distinct from the sound it replaced (cf. Zelinsky, 2003). But the acoustic manipulation had no effect on object-encoding performance, even though it resulted in more spectral differences within one of the scenes. Gregg and Samuel (2008) suggested that successful change detection may not be based on object identification, as has traditionally been assumed to underlie visual scene perception (e.g., Biederman, 1987; Edelman, 1998; Ullman, 2007), but is, instead, accomplished by comparing global acoustic representations of the scenes (similar claims have been offered for vision by Greene & Oliva, 2009; Oliva & Torralba, 2001). An example of such a global representation would be the overall distribution of energy at different frequencies, a feature that people use in identifying the timbre of individual musical instruments (Grey, 1977). This conclusion seems consistent with the reverse hierarchy theory's account of perceptual processing, which claims that perception of scenes generally takes place first at a high, categorical level of representation (Hochstein & Ahissar, 2002). Only when the task affords access to lower-level sensory details are such details available for perception. This makes the clear prediction that change detection failures should be more prevalent when the categories of objects are not changed in a scene despite the presence of large acoustic changes, which is consistent with results to be discussed shortly.

As was mentioned above, the suggestion by Gregg and Samuel (2008) that change deafness does not necessarily involve deficient object identification was based on an analysis of overall object-encoding and change detection performance, not a trial-by-trial analysis of whether poor encoding is accompanied by poor change detection. Recently, McAnally et al. (2010) did distinguish between object encoding on detected and not-detected change trials in audition. They found that performance in identifying which object was deleted was near ceiling when changes were detected but at chance level when changes were *not* detected. This finding suggests that changes may be detected only if objects are well encoded (contrary to the findings of Gregg & Samuel, 2008, and Mitroff, et al., 2004). However, it should be noted that the extent of change deafness that occurred in McAnally et al. was quite modest. They obtained 15% change deafness for scene sizes of four objects, whereas Gregg and Samuel (2008) obtained 45% change deafness for scene sizes of four

Fig. 1 Paradigm used in Mitroff et al. (2004). Figure reprinted with permission. Copyright 2004 by the Psychonomic Society

Fig. 2 Change detection and object encoding results in Gregg and Samuel (2008, Experiment 5): Percentages of errors on object -encoding and change detection questions as a function of the acoustics of the scene. 0 = neither frequency nor hamonicity changed from scene 1 to scene 2 (hardest condition based on acoustics); F1 = frequency change

from scene 1 to scene 2; H1 = harmonicity shift from scene 1 to scene 2; 2 = both frequency and harmonicity changed from scene 1 to scene 2 (easiest condition based on acoustics). Error bars represent the standard errors of the means. Figure reprinted with permission. Copyright 2008 by the American Psychological Association

objects. In McAnally et al., a changed scene consisted of an object that was missing, rather than an object replaced by a different object, as in Gregg and Samuel (2008).

In summary, the finding that objects appear to be well encoded even when changes to these objects are missed (Gregg & Samuel, 2008; Mitroff et al., 2004; but see McAnally et al., 2010) suggests that change blindness/deafness may not be caused by a faulty short-term memory system but, instead, may be caused by a faulty comparison process that may or may not depend on short-term memory (e.g., Angelone, Levin & Simons, 2003; Hollingworth & Henderson, 2002; Mitroff et al., 2004; Ryan & Cohen, 2004). Unfortunately, there is little or no direct evidence supporting this possibility, but this is an issue that warrants attention in future studies.

Another similarity between change blindness and change deafness is the amount and type of detail that appears to be extracted from scenes. In the auditory domain, Gregg and Samuel (2009) have shown that abstract identity information seems to be encoded preferentially, as compared with intricate physical detail, consistent with the reverse hierarchy theory (Hochstein & Ahissar, 2002). In this experiment, within-category changes (e.g., a large dog barking changing to a small dog barking) were missed more often than between-category changes (e.g., a large dog barking changing to a piano tune). It is important to note that this result occurred even though acoustic distance for within- and between-category changes was controlled for (see Fig. 3 for an acoustic distribution of the stimuli used in this experiment). In fact, the finding that within-category changes elicited more change deafness was so robust that it occurred even when the within-category changes were acoustically advantaged, as compared with between-category changes (i.e., within-category changes should have been easier to detect on the basis of the acoustics of

the scene). Similarly, Hollingworth and Henderson (2002) found that participants missed within-category changes to visual scenes (e.g., a spiral-edged notebook changing to a flat-edged notebook) more often than between-category changes (e.g., a spiral-edged notebook changing to a pen). Thus, in both vision and hearing, semantic information dominates sensory representations in memory, although physical features are likely to be accessed in the absence of semantic differences. These findings of semantic influences across modalities coincide well with the adaptive utility of detecting meaningful changes in real-world situations. However, they do not address the specific nature of the high-level representation being used. For example, it is possible that observers simply form a mental list of verbal labels for all of the objects in the prechange scene, as has

Fig. 3 Acoustic distribution of stimuli used in Gregg and Samuel (2009, Experiment 1; doi:10.3758/APP.71.3.607). The distance between within-category changes (in this figure, the change from dog 2 to dog 1) was matched for the distance between between-category changes (in this figure, the change from dog 2 to ship 2). Figure reprinted with permission. Copyright 2009 by the Psychonomic Society

been suggested (Demany, Trost, Serman & Semal, 2008). Alternatively, higher-order representations might be activated that reflects the similarity between objects within and between categories, such as in exemplar- or prototype-based models of visual object perception (Palmeri & Gauthier, 2004).

Differences Although there are clearly some similarities between change blindness and change deafness, several studies have indicated that there may be properties unique to change deafness, which indicates that the type of memory involved in change detection tasks may be different across modalities. For example, spatially separating auditory stimuli does not lead to improved detection (Gregg & Samuel, 2008), whereas cuing spatial location in vision improves detection (e.g., Stolz & Jolicoeur, 2004). There is also evidence for a relative lack of potency of spatial cues in auditory scene analysis tasks (e.g., McDonald & Alain, 2005). This difference across modalities is likely due to differences in physiological sensitivity to spatial information. The efficacy of spatial cues in change blindness is consistent with the visual system's inherently spatial (retinotopic) organization, whereas the primary organizational principle for auditory information is based on frequency (tonotopic; Kaas & Hackett, 2000; Merzenich & Brugge, 1973). The tonotopic organization of the auditory system predicts that separating auditory stimuli on the basis of spectral cues should enhance performance, and this is in fact the case (Gregg & Samuel, 2008). Thus, it appears that some differences between change blindness and change deafness may reflect the differences in the basic sensory encoding used in the two modalities, rather than differences in the mechanisms involved in auditory and visual change detection.

A second potential difference between change deafness and change blindness is that the presence of an interruption seems to be required to elicit change blindness, but not change deafness. In vision, change blindness presumably occurs when the luminance or color transient that would attract attention to the change is masked by an interruption, such as a blink (O'Regan, Deubel, Clark & Rensink, 2000) or a mudsplash (O'Regan, Rensink, & Clark, 1999). When the transient is not present, visual changes become quite easy to detect (Rensink et al., 1997). In audition, transients would typically be changes in energy or frequency from one scene to the next when a sound is removed or replaced. Like change blindness, change deafness occurs when an interruption between scenes masks the transients (e.g., Eramudugolla et al., 2005); however, change deafness is just as prevalent when there is no interruption between scenes (Pavani & Turatto, 2008). This finding seems to suggest that the auditory system does not rely on transients in the same manner as the visual system and, therefore, that auditory and visual memory during natural scene percep-

tion may operate quite differently. This conclusion is not necessarily correct, however, given the body of evidence that suggests that the auditory system is quite good at detecting transients (e.g., Coath & Denham, 2007). Furthermore, as Pavani and Turatto pointed out, the sounds they used were continuously transient. As a result, listeners in Pavani and Turatto's experiment may have missed the changes because the simultaneous transients occurring within the other sounds masked the transient occurring at the moment of the change. This explanation could be tested in the visual domain to determine whether visual stimuli with overlapping transients render the blank screen between scenes unnecessary. Or in the auditory domain, change detection could be investigated further using static auditory sounds (e.g., Demany et al., 2008).

Another difference between change blindness and change deafness is that the amount of time separating scene 1 and scene 2 has been found to affect the two phenomena differently. In vision, short-term memory is known to have a large capacity at short durations (less than 100 ms) that decreases at longer durations (Alvarez & Cavanagh, 2004; Luck & Vogel, 1997; Phillips, 1974). In audition, however, Demany et al. (2008) found that although an increase in the number of tones within a chord resulted in more difficulty in detecting frequency changes to one of the tones (for an illustration of the stimuli, see Fig. 4), a similar decline in performance with longer time delays regardless of the number of tones present suggested an unlimited capacity for auditory memory and a lack of change deafness. Demany et al. (2008) concluded that reports of change deafness are not analogous to reports of change blindness and actually reflect errors in verbal memory. There are a few problems with Demany et al.'s (2008) claims, however. First, the notion of an unlimited capacity auditory memory

Fig. 4 Depiction of stimuli and tasks used in Demany et al. (2008). Horizontal dashes represent pure tones. The correct response to the trial on the left would be *different*, and the correct response to the trial on the right would be *downward*. Figure reprinted with permission. Copyright 2008 by the Association for Psychological Science

system seems implausible. Furthermore, memory for frequency or frequency changes may be specialized (Demany, Pressnitzer & Semal, 2009; Demany & Ramos, 2005; Demany, Semal & Pressnitzer, 2011), meaning that while some specific types of changes (e.g., frequency changes to static tones) are detected with ease, other changes might still be difficult to detect. Thus, while the existence of automatic change detection is compatible with the phenomenon of change deafness for complex objects, the former process may be limited to individual pure tones that change in frequency. Also, one could argue that the perceptually fused tone components within chords are not analogous to objects in a visual scene and may not be considered “objects” by the auditory system.

Summary Change blindness and change deafness seem to have more similarities than differences. Aside from differences in input codes, the processes involved in change detection may be general mechanisms, rather than ones specific to a modality. This raises the need for future studies that determine the extent to which similar mechanisms in different modality-specific neural circuits are involved in auditory and visual change detection, as opposed to the possibility that some aspects of change detection actually rely on the same underlying brain regions. Future research would also benefit from a focus on resolving theoretical issues that remain about why failures to detect changes occur. For example, the issue still remains as to what extent sensory encoding, attention, or comparison processes are responsible for failures to detect changes and how the interaction of these processes with auditory memory contributes to change deafness.

Event-related potential and functional imaging studies

One area of research that is particularly promising for investigating the processes involved in both change blindness and change deafness is the growing body of recent neurophysiological investigations. For example, several researchers have found evidence for an automatic change detection process. Busch, Fründ, and Herrmann (2010) found that visual changes can be detected even when participants cannot identify the object that changes. Event-related potentials (ERPs) from this study indicated that detection of change and identification of the object that changed may rely on different neural processes. A similar study demonstrated that changes can be accurately localized in space without being accurately identified. Specifically, ERPs indicated that localization and identification of changes rely on the same early, sensory-level neural processes but differ at later stages of neural processing (Busch, Dürschmid & Herrmann, 2010). Other research has found neural markers of successful change detection, such as the N2pc component (e.g., Eimer & Mazza, 2005;

Schankin & Wascher, 2008). The visual N2pc response likely reflects enhanced attention to the changing object (Eimer & Mazza, 2005). The N2ac is an auditory analogue to the N2pc that could be used to study change deafness (Gamble & Luck, 2011). The mismatch negativity might also be another useful ERP component for studying change deafness, because it is thought to reflect the detection of change with respect to information stored in auditory sensory memory, even when that information is somewhat abstract (Näätänen & Winkler, 1999). Early visual sensory activity has also been found to be enhanced for successfully detected changes, which indicates that there may be an enhanced orienting to the basic sensory attributes of a scene that facilitates change detection (Eimer & Mazza, 2005; Pourtois, De Preto, Hauert & Vuilleumier, 2006). Brain imaging and transcranial magnetic stimulation studies have indicated an important role of the right visual dorsal pathway in successful change detection (Beck, Muggleton, Walsh & Lavie, 2006; Beck, Rees, Frith & Lavie, 2001). And a multimodal imaging experiment has found that the right temporoparietal junction, inferior frontal gyrus, and insula areas are important for detecting changes in the visual, auditory, and tactile modalities (Downar, Crawley, Mikulis & Davis, 2000).

There is no published neuroimaging or ERP work on change deafness that we know of, but there are many theoretical issues that would benefit from the insight such work could provide. For example, finding a neural marker of successful change detection in change deafness paradigms would help to determine whether the markers that have been found in change blindness experiments reflect a general process, rather than neural activity specific to visual stimuli. An investigation of automatic auditory change detection for complex scenes would also inform this issue. Finally, examination of the neural responses to the pre- and postchange scenes and a comparison of those markers with object-encoding ability could help to determine the sensory and cognitive stages of processing involved in change deafness.

Recognition and recall of sounds

Long-term memory

As with change detection, it is also informative to compare recognition and recall across modalities to characterize abilities and to determine whether similar mechanisms are at work. A recent study by Cohen, Horowitz and Wolfe (2009) presented evidence that visual recognition memory of real-world objects is superior to auditory memory for the same objects. Participants were instructed to memorize sets of 64 or 90 natural sounds (e.g., birds chirping, a coffee

shop, popular music excerpts, spoken language excerpts). Immediately after trying to memorize the sounds, the participants were presented with another set of sounds, consisting of half *new* sounds and half *old* sounds that were in the original set they had tried to memorize. The task was simply to identify new versus old sounds. Performance was not as good ($d = 1.68 - 2.70$) as the remarkable ability of participants to memorize and later recognize static pictures ($d = 3.57$) that contained the same types of objects that produced the sounds (see Fig. 5). Even presenting the pictures or a verbal description along with the sounds during the memorization period did not enable auditory recognition to approach visual performance. The best auditory recognition was when spoken language excerpts were used, but recognition performance was still appreciably lower than during visual recognition. Music excerpts were remembered even worse than auditory objects. The only way it was possible to equate recognition performance between auditory and visual stimuli was by degrading the images, making the visual stimuli much less recognizable than the auditory stimuli. Other studies have documented the remarkable ability of human observers to recognize previously viewed visual pictures at rates greater than 90% correct, even when set sizes are much larger than those in Cohen et al. (2009; e.g., Standing, 1973). Thus, the most straightforward conclusion that can be drawn from these observations is that the ability to encode auditory objects and later recognize them is inferior to encoding and recognizing visual objects.

However, it is important to consider other explanations. It is possible that we have more experience attentively viewing visual objects and scenes and that this results in more finely tuned neural resources being available for encoding, storing, or retrieving visual scenes. One way to evaluate the role of experience is by testing individuals with exceptional auditory experience. Cohen and colleagues

have indeed shown that expert musicians are better than non-musically trained individuals at recognizing auditory objects, although the musicians were still poorer at recognizing auditory objects, as compared with visual objects (Cohen, Evans, Horowitz & Wolfe, 2011). It is possible that despite spending large amounts of time attending to sound, musicians still spend more time attending to visual scenes. Thus, testing congenitally blind individuals would be interesting, in order to see whether their ability to recognize auditory objects approaches the ability of sighted individuals to recognize visual objects. It would also be worthwhile testing young infants who do not have as much experience with visual scenes (or who may attend to sound to a larger extent than do adults) to see whether they fail to show a visual advantage for remembering objects. This would be especially interesting given that infants hear sounds in utero, prior to visual experience with the corresponding objects (DeCasper & Fifer, 1980).

In addition to the study by Cohen et al. (2011), another recent study examined auditory long-term memory (Crutcher & Beer, 2011)—in this case, to test whether an auditory analogue to the picture superiority effect (the superior ability to recognize pictures, as compared with visually presented words, Paivio & Csapo, 1973) exists. Their experiments showed superior auditory free recall (i.e., writing down each word they could remember having been presented previously) of object sounds, as compared with aurally presented words. Across all of the experiments, participants showed superior recall of natural nonspeech sounds, regardless of whether stimulus type was a within- or a between-subjects factor, whether participants were aware that their memory for the stimuli would be tested later, and whether they had a secondary task or not during memory encoding (see Fig. 6). The only case in which memory for words was similar to memory for natural nonspeech sounds was when participants were encouraged

Fig. 5 Recognition memory results as measured by d' from Cohen et al. (2009). Error bars represent the standard errors of the means. Experiment 1 tested memory for studied environmental sounds only. Experiment 2 compared memory for studied environmental sounds alone, environmental sounds presented in combination with pictures

of the sounds, environmental sounds presented in combination with the typed names of the sounds, the names alone, and the pictures alone. Experiment 3 compared memory for studied excerpts of music with studied excerpts of spoken language. Figure reprinted with permission. Copyright 2009 by the National Academy of Sciences

Fig. 6 Results from Crutcher and Beer (2011, Experiment 4; doi:10.3758/s13421-010-0015-6): Recall as a function of stimulus type (the sounds of objects vs. spoken verbal label of the objects) and encoding condition (incidental = during study, participants were told to write down the name of each sound but were not told that they would have to recall them later; intentional = participants were told to write down the sounds and that they would be tested later; free strategy = participants were told to do whatever they could to remember each studied item but did not write down the names). Figure reprinted with permission. Copyright 2011 by the Psychonomic Society

to form a mental image of the words (i.e., to imagine the sound to which a word referred). This effect of imagery supported the idea that memory for nonspeech sounds is superior to memory for words because nonspeech sounds are encoded pictorially and verbally, whereas words are encoded only in verbal format, unless specific instructions are given to form mental images. The fact that an auditory equivalent of picture superiority occurs suggests, again, that similar memory mechanisms exist in hearing and vision. However, further research on auditory picture superiority is clearly warranted to determine whether it occurs in other circumstances, such as in recognition tasks or with larger sets of stimuli that require memorizing.

Short-term memory

In interpreting the findings of Cohen et al. (2009), it is important to consider that the auditory stimuli were dynamic sounds with continuous spectro-temporal changes, whereas the visual stimuli were static pictures. This naturally raises the issue of whether dynamic stimuli are more difficult to remember than static stimuli, regardless of differences in sensory modality. A related issue is whether dynamic stimuli are better remembered when presented in one modality, as compared with the other modality. Previous studies comparing short-term auditory and visual memory have shown that there is a memory advantage for sequential auditory stimuli (as compared with sequential

visual stimuli), such as when people try to recognize or recall verbal material, musical material, or binary sequences (e.g., Crowder, 1986; Degelder & Vroomen, 1992; Duis, Dean & Derks 1994; Henmon, 1912; McFarland & Cacace, 1995; Roberts, 1986; Schendel & Palmer, 2007; for theoretical treatments, see Crowder & Morton, 1969; Glenberg & Swanson, 1986; Penney, 1989; for a comprehensive review of the older literature, see Penney, 1975). This well-studied auditory advantage, often called the *modality effect*, is often the result of a greater recency effect for auditory stimuli, but some studies have shown a more general auditory advantage throughout comparable lists from the two modalities (e.g., spoken vs. visually presented digits).

While these studies on the modality effect typically have assessed short-term memory, the classic study on long-term memory for visual scenes by Standing (1973) showed that recognition of visually presented words, aurally presented words, and aurally presented music was roughly the same, although all of these conditions resulted in worse performance than did the one with visual scenes. Thus, including memory for sequentially presented material in short-term memory, there are at least two examples in which auditory memory is comparable or superior to visual memory. So there appears to be no strong reason to believe that memory for visual material is inherently superior. Rather, there may be a visual advantage for static scenes, recognizable environmental objects, or other stimuli that present large amounts of simultaneous information. Determining whether dynamic visual scenes (e.g., a video of a crowded café scene) are better remembered than corresponding dynamic auditory scenes (e.g., the sound track of a café scene) requires further study.

Any advantage that visual memory has for large amounts of simultaneous information may be limited to long-term memory. It is thought that short-term visual memory is severely limited by the number of objects (Cowan, 2010; Luck & Vogel, 1997) and also the amount of information contained in the limited number of objects (Alvarez & Cavanagh, 2004). A recent study, discussed earlier, suggested that such a limit may not exist in auditory short-term memory (Demany et al., 2008), when participants were asked to indicate whether a pure-tone component embedded in a chord consisting of other pure tones moved up or down in frequency after a delay of 0–2,000 ms. Although performance was better when chords consisted of fewer pure-tone components (e.g., 4 vs. 7 or 12), the observed decline in memory with longer interstimulus intervals was similar regardless of the scene size, in contrast to visual memory, which declines more rapidly for more complex visual scenes (Phillips, 1974). This suggested that while scene size affects sensory and/or attentional factors that are important for discrimination, scene size does not affect the

ability to store information in auditory short-term memory, unlike visual memory. An additional recent study by the same laboratory directly compared auditory memory for the frequency of tones in a chord and visual memory for the location of dots in a circular array, finding superior performance in the visual than in the auditory modality when the task was to judge whether a test element (i.e., a tone or a dot) was present or absent in the previous sample array of elements (Demany, Semal, Cazalets & Pressnitzer 2010). However, superior performance was observed in the auditory, as compared with the visual, modality when the task was to judge the direction of movement of an element from the sample to the test.

An important limitation of these findings by Demany and colleagues is that they may be specific to memory for items that can be encoded in terms of static frequency, given that studies of auditory memory have found evidence for different memory systems for different auditory perceptual attributes, such as pitch, timbre, and loudness (Clement, Demany & Semal, 1999; McKeown & Wellsted, 2009; Mercer & McKeown, 2010; Semal & Demany, 1991). Along these lines, there appear to be specialized frequency-shift detectors that may aid discrimination without engaging normal auditory short-term memory mechanisms (Cousineau, Demany & Pressnitzer, 2009; Demany et al., 2009; Demany & Ramos, 2005; Demany et al., 2011). It might be expected that memory for frequency is somehow special, given that it is a major organizing dimension in the auditory system in the form of tonotopic maps up to the level of the primary auditory cortex (Kaas & Hackett, 2000; Merzenich & Brugge, 1973). However, this does not appear to be the case in visual short-term memory for spatial patterns, despite their privileged retinotopic encoding in the visual system, since only very short-lasting iconic memory is thought to be unlimited in capacity (Coltheart, 1980; Phillips, 1974; Sperling, 1960). Finally, it is important to also point out that even though the visual and auditory stimuli in Demany et al. (2010) were qualitatively similar, it is difficult to be sure that the information load was quantitatively the same across modalities. Future studies might therefore attempt to quantitatively equate information load (e.g., Alvarez & Cavanagh, 2004). However, it would also be important to use methods established in the visual domain (e.g., Alvarez & Cavanagh, 2004; Luck & Vogel, 1997; Phillips, 1974) to study whether the maximum number of objects or pieces of information that can be held in memory is similar across modalities, which has only begun to be directly addressed by recent studies (Demany et al., 2010; Demany et al., 2008).

Several studies have provided evidence that auditory short-term memory and visual short-term memory rely on similar mechanisms, which at least qualifies the claims by Demany and colleagues. For example, recognition of binary sequences (i.e., patterns of low and high tones or patterns of

blue and yellow lights) resulted in very similarly shaped serial position memory functions and similar temporal decay functions for auditory and visual conditions, despite overall better performance for auditory sequences (McFarland & Cacace, 1995). In a study comparing articulatory suppression effects (i.e., impairment of memory for study material due to vocally producing irrelevant sound), a similar disruption of memory occurred for verbal (lists of numbers) and musical (melodic sequences) material, whether the material was presented in the visual (printed numbers or notes on a musical staff) or auditory (spoken numbers or tones) modality (Schendel & Palmer, 2007). Furthermore, visual distraction stimuli (black-and-white checkerboard grids) interfered only with memory for visual note sequences, unless participants were encouraged to translate auditory stimuli into a visual code, in which case the visual distraction impaired memory for all the stimuli. This study suggests that at least for verbal and musical materials, auditory and visual stimuli can flexibly engage the same memory mechanisms. Using sequential presentations of visually presented unfamiliar faces and aurally presented non-words, similar serial position curves were found (Ward, Avons & Melling, 2005). Finally, a recent study compared memory for sequentially presented nonmeaningful auditory (moving sinusoidal ripples) and visual (moving and stationary sinusoidal gratings) stimuli (Visscher, Kaplan, Kahana & Sekuler, 2007). Although the researchers found better overall recognition for auditory stimuli, they found similar effects of list length, retention interval, and serial position for all stimulus types. These findings require replication, however, because some of the interactions between these factors and stimulus modality approached significance, suggesting some quantitative (if not qualitative) modality differences.

A challenge that has been highlighted above is how to study memory for complex patterns typical of real-world situations, while at the same time having enough experimental power to distinguish between memory for the acoustics of a sound versus its verbal or semantic characteristics. In addition to the study by Visscher et al. (2007), another recent study offers a promising approach to this issue by examining the effect of prior exposure to acoustic Gaussian noise patterns on the ability to discriminate the same noises (Agus, Thorpe & Pressnitzer, 2010). Participants had to detect whether noise patterns contained two exact repetitions of the spectro-temporal energy within a 1-s sample from noise patterns that had no such repetition within the same sample duration. On trials containing repeating noises that had been presented on previous trials, detection performance was enhanced, as compared with when the repeated noise had not been presented previously, demonstrating an implicit memory for the noises. Although it is not clear whether the stimuli from this study could be

used for studies of more explicit forms of memory, such as in the recognition and recall experiments described thus far, it is certainly worth exploring.

Summary Researchers have begun to raise some interesting questions about the abilities and mechanisms underlying auditory long-term and short-term memory. While there is some suggestion that long-term memory for objects or scenes may be superior in the visual domain, this finding requires independent replication using a wider variety of stimulus materials and tasks to determine how general the finding is. In short-term memory, it is less clear that one modality is advantaged over the other, and the evidence thus far suggests that it depends on the stimuli used and the type of task required of participants.

Event-related potential and functional imaging studies

A small number of studies measuring brain activity have provided further information regarding the similarities and differences between auditory and visual memory. Some of these studies compared auditory and visual verbal memory, although there are also data on memory for recognizable nonverbal objects. One study used ERPs to study recognition of digits (1 through 9) presented in lists with lengths of one, three, or five items (Pratt, Michalewski, Patterson & Starr, 1989). During the recognition test after each list, a probe was presented, with half of the probes being *old* items and half *new* (i.e., previously presented in the list). Behavioral recognition performance of the probes was at ceiling for both auditory and visual digits, although reaction time was significantly longer with the larger set sizes for both modalities. For both modalities, the probe elicited a frontal P3a response and a parietal P3b response, which may index attention-orienting and memory-updating functions, respectively (Polich, 2007). The P3a and P3b were generally larger in amplitude and earlier in latency for visual probes, as compared with auditory probes. In both modalities, there were also set size effects, with larger amplitudes and earlier latencies for smaller set sizes. This study therefore suggested that similar neural processes are activated during auditory and visual recognition.

In an ERP study of two auditory and visual working memory tasks for nonmeaningful noisy band-pass stimuli, three stimuli were presented sequentially, and participants were asked to judge whether (1) the highest frequency stimulus was presented first, second, or third or (2) the third stimulus was lower, intermediate, or higher in frequency than the first two stimuli (Protzner, Cortese, Alain & McIntosh, 2009). For both tasks, the results showed a large degree of modality specificity in early sensory-related activity immediately following stimulus presentations, as would be expected, but similar P3b responses for both modalities, as in the study by Pratt et al. (1989).

Additional evidence for similar memory mechanisms across modalities is found in a study comparing memory for visual and auditory words (Kayser, Fong, Tenke & Bruder, 2003). This study used a continuous recognition paradigm, in which a long series of stimuli were presented, some of which were old and some of which were new. This study found a frontal late negative response that was larger for correctly identified old than for correctly identified new items, an effect that has been attributed to a familiarity-based recognition process (Rugg & Curran, 2007). An additional smaller positive enhancement for old items was observed, which may reflect recollection-based recognition (Rugg & Curran, 2007). Despite some differences in which electrodes manifested the old versus new effects and latency differences across modalities, the pattern of results was remarkably similar during visual and auditory recognition. Another ERP study using auditory and visual words presented in separate study blocks, followed by visual-only test blocks (half old and half new words) provided further information regarding the degree of modality specificity in recognition memory (Curran & Dien, 2003). This study observed a frontal late negativity and a parietal late positivity for old words that occurred in both modalities, as in the study by Kayser et al. However, an earlier positivity for old words occurred only for stimuli studied in the visual modality, suggesting a priming mechanism that was modality specific. Thus, highly similar mechanisms of recognition memory seem to be present across modalities, although there is clearly the possibility that at least partially distinct modality-specific neural generators are active.

To our knowledge, Chao and Knight (1996); Chao, Nielsen-Bohman & Knight (1995) performed the only ERP studies testing auditory memory of environmental sounds. In both studies, they found larger and earlier parietal P3b and slow wave responses for old stimuli repeated after a short delay, as compared with new stimuli and old stimuli presented after a long delay. A frontal N4 response, associated with semantic encoding in both audition and vision (Kutas & Hillyard, 1984), was especially prominent for initial presentation of stimuli and for old stimuli presented after a long delay. The N4 did not differ between stimuli that were verbally encodable versus those that were not, suggesting that this response indexes general semantic encoding, as opposed to language-specific encoding. The authors interpreted the P3b response as reflecting the robustness of template matching of prior stimuli with a current stimulus, consistent with the larger response for recent old items that result in better recognition performance. Although this study did not directly compare the results with visual recognition, the similarity is striking with other studies discussed above—in particular, those reporting a parietal positive response that is larger for old items and is likely associated with recollection memory mechanisms.

A study using functional magnetic resonance imaging during continuous word recognition provided more detailed anatomical evidence about the cortical areas involved in auditory and visual memory (Buchsbaum, Padmanabhan & Berman 2010). Across both modalities, old words resulted in enhanced activity, as compared with new words, in a number of brain areas (intraparietal sulcus, left anterior prefrontal cortex, anterior insula, and dorsal precentral sulcus), while enhanced activity for new words occurred in distinct areas (parahippocampal gyrus, ventral occipital cortex, medial fronto-polar cortex, and anterior-superior temporal sulcus). Brain areas specifically responsive to the duration of time between the first and repeated presentations of old items included the inferior frontal, inferior parietal, and middle and superior temporal cortices. This included areas that increased their activity with longer lags, those that decreased with longer lags, and those that initially showed repetition suppression followed by increased activity with longer lags. Better memory performance in individuals correlated with less old > new activity in left-hemisphere regions such as the superior temporal sulcus, inferior frontal gyrus, ventral temporal cortex, and anterior hippocampus and with more old > new activity in the superior frontal gyrus, supramarginal gyrus, and left anterior cingulate gyrus. Although explicit comparisons were not made across modalities, many of the brain areas activated are largely those implicated in modality-general cognitive functions. Fewer of the reported brain changes were in occipital and superior temporal areas that are more traditionally associated with visual and auditory processing, respectively. In summary, neurophysiological studies of recognition and recall memory suggest many similarities in the processes underlying auditory and visual memory. However, in many cases, it is not clear whether the similarities occur at all stages of processing (e.g., encoding and retrieval). Also, the extent to which similar processes occur in modality-specific neural circuits, as opposed to the two modalities actually sharing neural resources, is unclear.

Conclusions

The goal of this review has been to discuss some of the key issues in the study of auditory memory and how recent research has addressed these issues and to provide suggestions for future research in this area. As was noted in the introduction, one of the most important issues in auditory memory research is determining the abilities and limitations of auditory memory, what neural mechanisms account for observed patterns of behavior, and to what extent auditory memory differs from visual memory. We reviewed several different lines of research that have been particularly informative about these issues, including

research on change detection, auditory long-term memory, and auditory short-term memory. The majority of the results support the conclusion that auditory and visual memories have more similarities than differences. Aside from differences in input codes, the processes underlying particular types of auditory and visual memory may be general, rather than modality specific. However, much additional work is required to develop a more advanced understanding of auditory memory and its similarity to memory for stimuli in other modalities. Similarly, it is important to understand how memory for different types of auditory stimuli (e.g., speech vs. music, animate vs. inanimate) relies on the same or different mechanisms, in addition to how memory differs for different low-level stimulus features (e.g., Semal & Demany, 1991).

Future studies are needed to examine the extent to which similar mechanisms in different modality-specific neural circuits are involved in auditory and visual memory. For example, it remains a possibility that some of the similarities in memory performance across modalities are due to the utilization of common verbal or semantic memory mechanisms regardless of the input modality. On the other hand, it is also important to consider theoretical and empirical work suggesting that verbal and semantic representations are not as distinct from sensory-motor representations (Barsalou, Simmons, Barbey & Wilson, 2003; Glenberg, 1997; Pulvermuller, 2005), as in more traditional modular accounts of cognition (Fodor, 1983). Thus, it is possible that even verbal and semantic representations are implemented in sensory-processing brain regions that are most relevant to the content of what is being represented, effectively blurring the distinction between sensory and verbal or semantic representations. Such “high-level” processing in anatomically low-level brain areas might occur through the reactivation of neural circuits that were initially involved in processing physically presented stimuli (Barsalou et al., 2003). Additional studies are also needed to reveal the specific transformations of auditory information into memory traces, how these memory traces are accessed, how they are compared with incoming sensory input, and the extent to which these operations depend on the types of sounds being processed (e.g., meaningful vs. nonmeaningful, familiar vs. unfamiliar, environmental vs. speech vs. music). Finally, future studies should continue uncovering new knowledge and new methods that will aid in comparing memory mechanisms across modalities, while controlling as much as possible for differences in how stimuli are processed and for differences in storage capacity.

While a great amount of information has been gathered using behavioral measurements, it seems likely that further insights will arise from a more diverse set of approaches, including further behavioral studies, neurophysiological

measurements, and correlations with particular anatomical features such as gray and white matter volumes and white matter tract structure. In addition to these correlation-based approaches, more causal forms of evidence, such as brain stimulation that disrupts human performance, examination of patients with gross brain lesions or with disorders that are known to affect the auditory system, and animal models, could also prove to be useful. Finally, as more is known about the nature of auditory memory, the development of computational models will likely be important to test ideas about the particular mechanisms involved in encoding auditory information, detecting changes, and remembering previously presented sounds.

Author Note The authors thank Simon Grondin, Mari Jones, Arty Samuel, David Weintraub, Erin Hannon, and anonymous reviewers for their helpful comments.

References

- Agus, T. R., Thorpe, S. J., & Pressnitzer, D. (2010). Rapid formation of robust auditory memories: Insights from noise. *Neuron*, 66, 610–618.
- Agres, K. R., & Krumhansl, C. L. (2008). Musical change deafness: The inability to detect change in a non-speech auditory domain. In B. C. Love, K. McRae, & V. M. Sloutsky (Eds.), *Proceedings of the 30th Annual Conference of the Cognitive Science Society* (pp. 969–974). Austin, TX: Cognitive Science Society.
- Alain, C. (2007). Breaking the wave: Effects of attention and learning on concurrent sound perception. *Hearing Research*, 229, 225–236.
- Alain, C., & Arnott, S. R. (2000). Selectively attending to auditory objects. *Frontiers in Bioscience*, 5, D202–D212.
- Alain, C., Dyson, B. J., & Snyder, J. S. (2006). Aging and the perceptual organization of sounds: A change of scene? In P. M. Conn (Ed.), *Handbook of models for human aging* (pp. 759–769). Amsterdam: Elsevier.
- Alvarez, G. A., & Cavanagh, P. (2004). The capacity of visual short-term memory is set both by visual information load and by number of objects. *Psychological Science*, 15, 106–111.
- Angelone, B. L., Levin, D. T., & Simons, D. J. (2003). The roles of representation and comparison failures in change blindness. *Perception*, 32, 947–962.
- Auvray, M., Gallace, A., Hartcher-O'Brien, J., Tan, H. Z., & Spence, C. (2008). Tactile and visual distractors induce change blindness for tactile stimuli presented on the fingertips. *Brain Research*, 1213, 111–119.
- Barsalou, L. W., Simmons, W. K., Barbey, A. K., & Wilson, C. D. (2003). Grounding conceptual knowledge in modality-specific systems. *Trends in Cognitive Sciences*, 7, 84–91.
- Beck, D. M., Muggleton, N., Walsh, V., & Lavie, N. (2006). Right parietal cortex plays a critical role in change blindness. *Cerebral Cortex*, 16, 712–717.
- Beck, D. M., Rees, G., Frith, C. D., & Lavie, N. (2001). Neural correlates of change detection and change blindness. *Nature Neuroscience*, 4, 645–650.
- Biederman, I. (1987). Recognition-by-components: A theory of human image understanding. *Psychological Review*, 94, 115–147.
- Bregman, A. S. (1990). *Auditory scene analysis: The perceptual organization of sound*. Cambridge, MA: MIT Press.
- Buchsbaum, B. R., Padmanabhan, A., & Berman, K. F. (2010). The neural substrates of recognition memory for verbal information: Spanning the divide between short- and long-term memory. *Journal of Cognitive Neuroscience*, 23, 978–991.
- Busch, N. A., Dürschmid, S., & Herrmann, C. S. (2010). ERP effects of change localization, change identification, and change blindness. *NeuroReport*, 31, 371–375.
- Busch, N. A., Fründ, I., & Herrmann, C. S. (2010). Electrophysiological evidence for different types of change detection and change blindness. *Journal of Cognitive Neuroscience*, 22, 1852–1869.
- Carlyon, R. P. (2004). How the brain separates sounds. *Trends in Cognitive Sciences*, 8, 465–471.
- Chao, L. L., & Knight, R. T. (1996). Prefrontal and posterior cortical activation during auditory working memory. *Cognitive Brain Research*, 4, 27–37.
- Chao, L. L., Nielsen-Bohlman, L., & Knight, R. T. (1995). Auditory event-related potentials dissociate early and late memory processes. *Electroencephalography and Clinical Neurophysiology*, 96, 157–168.
- 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.
- Clement, S., Demany, L., & Semal, C. (1999). Memory for pitch versus memory for loudness. *Journal of the Acoustical Society of America*, 106, 2805–2811.
- Coath, M., & Denham, S. L. (2007). The role of transients in auditory processing. *Biosystems*, 89, 182–189.
- Cohen, M. A., Evans, K. K., Horowitz, T. S., & Wolfe, J. M. (2011). Auditory and visual memory in musicians and non-musicians. *Psychonomic Bulletin & Review*, 18, 586–591.
- Cohen, M. A., Horowitz, T. S., & Wolfe, J. M. (2009). Auditory recognition memory is inferior to visual recognition memory. *Proceedings of the National Academy of Sciences*, 106, 6008–6010.
- Coltheart, M. (1980). Iconic memory and visible persistence. *Perception & Psychophysics*, 27, 183–228.
- Committee on Hearing. (1988). Speech understanding and aging. *Journal of the Acoustical Society of America*, 83, 859–895.
- Cousineau, M., Demany, L., & Pressnitzer, D. (2009). What makes a melody: The perceptual singularity of pitch sequences. *Journal of the Acoustical Society of America*, 126, 3179–3187.
- Cowan, N. (1984). On short and long auditory stores. *Psychological Bulletin*, 96, 341–370.
- Cowan, N. (2010). The magical mystery four: How is working memory capacity limited, and why? *Current Directions in Psychological Science*, 19, 51–57.
- Crowder, R. G. (1986). Auditory and temporal factors in the modality effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 12, 268–278.
- Crowder, R. G., & Morton, J. (1969). Precategorical Acoustic Storage (PAS). *Perception & Psychophysics*, 5, 365–373.
- Crutcher, R. J., & Beer, J. M. (2011). An auditory analog of the picture superiority effect. *Memory & Cognition*, 39, 63–74.
- Curran, T., & Dien, J. (2003). Differentiating amodal familiarity from modality-specific memory processes: An ERP study. *Psychophysiology*, 40, 979–988.
- Dahmen, J. C., & King, A. J. (2007). Learning to hear: Plasticity of auditory cortical processing. *Current Opinion in Neurobiology*, 17, 456–464.
- Darwin, C. J. (1997). Auditory grouping. *Trends in Cognitive Sciences*, 1, 327–333.
- DeCasper, A. J., & Fifer, W. P. (1980). Of human bonding: Newborns prefer their mothers' voices. *Science*, 208, 1174–1176.
- Degelder, B., & Vroomen, J. (1992). Abstract versus modality-specific memory representations in processing auditory and visual speech. *Memory & Cognition*, 20, 533–538.

- Demany, L., Pressnitzer, D., & Semal, C. (2009). Tuning properties of the auditory frequency-shift detectors. *Journal of the Acoustical Society of America*, 126, 1342–1348.
- Demany, L., & Ramos, C. (2005). On the binding of successive sounds: Perceiving shifts in nonperceived pitches. *Journal of the Acoustical Society of America*, 117, 833–841.
- Demany, L., & Semal, C. (2008). The role of memory in auditory perception. In W. A. Yost, A. N. Popper, & R. R. Fay (Eds.), *Auditory perception of sound sources* (pp. 77–113). New York: Springer.
- Demany, L., Semal, C., Cazalets, J. R., & Pressnitzer, D. (2010). Fundamental differences in change detection between vision and audition. *Experimental Brain Research*, 203, 261–270.
- Demany, L., Semal, C., & Pressnitzer, D. (2011). Implicit versus explicit frequency comparisons: Two mechanisms of auditory change detection. *Journal of Experimental Psychology: Human Perception and Performance*, 37, 597–605.
- Demany, L., Trost, W., Serman, M., & Semal, C. (2008). Auditory change detection: Simple sounds are not memorized better than complex sounds. *Psychological Science*, 19, 85–91.
- DeWeese, M. R., Hromádka, T., & Zador, A. M. (2005). Reliability and representational bandwidth in the auditory cortex. *Neuron*, 48, 479–488.
- Downar, J., Crawley, A. P., Mikulis, D. J., & Davis, K. D. (2000). A multimodal cortical network for the detection of changes in the sensory environment. *Nature Neuroscience*, 3, 277–283.
- Duis, S. S., Dean, R. S., & Derks, P. (1994). The modality effect: A result of methodology? *International Journal of Neuroscience*, 78, 1–7.
- Dyson, B. J., & Ishfaq, F. (2008). Auditory memory can be object based. *Psychonomic Bulletin & Review*, 15, 409–412.
- Edelman, S. (1998). Representation is representation of similarities. *Behavioral and Brain Sciences*, 21, 449–467.
- Eimer, M., & Mazza, V. (2005). Electrophysiological correlates of change detection. *Psychophysiology*, 42, 328–342.
- Eramudugolla, R., Irvine, D. R. F., McAnally, K. I., Martin, R. L., & Mattingley, J. B. (2005). Directed attention eliminates 'change deafness' in complex auditory scenes. *Current Biology*, 15, 1108–1113.
- Feldman, J. (2003). What is a visual object? *Trends in Cognitive Sciences*, 7, 252–256.
- Fodor, J. A. (1983). *Modularity of mind*. Cambridge, MA: MIT Press.
- Gallace, A., Tan, H. Z., & Spence, C. (2006). The failure to detect tactile change: A tactile analogue of visual change blindness. *Psychonomic Bulletin & Review*, 13, 300–303.
- Gamble, M. L., & Luck, S. J. (2011). N2ac: An ERP component associated with the focusing of attention within an auditory scene. *Psychophysiology*, 48, 1057–1068.
- Glenberg, A. M. (1997). What memory is for. *Behavioral and Brain Sciences*, 20, 1–55.
- Glenberg, A. M., & Swanson, N. G. (1986). A temporal distinctiveness theory of recency and modality effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 12, 3–15.
- Greene, M. R., & Oliva, A. (2009). Recognition of natural scenes from global properties: Seeing the forest without representing the trees. *Cognitive Psychology*, 58, 137–176.
- Gregg, M. K., & Samuel, A. G. (2008). Change deafness and the organizational properties of sounds. *Journal of Experimental Psychology: Human Perception and Performance*, 34, 974–991.
- Gregg, M. K., & Samuel, A. G. (2009). The importance of semantics in auditory representations. *Attention, Perception, & Psychophysics*, 71, 607–619.
- Grey, J. M. (1977). Multidimensional perceptual scaling of musical timbres. *Journal of the Acoustical Society of America*, 61, 1270–1277.
- Griffiths, T. D., & Warren, J. D. (2004). What is an auditory object? *Nature Reviews Neuroscience*, 5, 887–892.
- Henmon, V. A. C. (1912). The relation between mode of presentation and retention. *Psychological Review*, 19, 79–96.
- Hochstein, S., & Ahissar, M. (2002). View from the top: Hierarchies and reverse hierarchies in the visual system. *Neuron*, 36, 791–804.
- Hollingworth, A., & Henderson, J. M. (2002). Accurate visual memory for previously attended objects in natural scenes. *Journal of Experimental Psychology: Human Perception and Performance*, 28, 113–136.
- Irwin, D. E., & Andrews, R. (1996). Integration and accumulation of information across saccadic eye movements. In T. Inui & J. L. McClelland (Eds.), *Attention and performance XVI: Information integration in perception and communication* (pp. 125–155). Cambridge, MA: MIT Press.
- Jaaskelainen, I. P., Ahveninen, J., Belliveau, J. W., Raij, T., & Sams, M. (2007). Short-term plasticity in auditory cognition. *Trends in Neurosciences*, 30, 653–661.
- Kaas, J. H., & Hackett, T. A. (2000). Subdivisions of auditory cortex and processing streams in primates. *Proceedings of the National Academy of Sciences*, 97, 11793–11799.
- Kayser, J., Fong, R., Tenke, C. E., & Bruder, G. E. (2003). Event-related brain potentials during auditory and visual word recognition memory tasks. *Cognitive Brain Research*, 16, 11–25.
- Kubovy, M., & Van Valkenburg, D. (2001). Auditory and visual objects. *Cognition*, 80, 97–126.
- Kutas, M., & Hillyard, S. A. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature*, 307, 161–163.
- Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature*, 390, 279–281.
- McAnally, K. I., Martin, R. L., Eramudugolla, R., Stuart, G. W., Irvine, D. R. F., & Mattingley, J. B. (2010). A dual-process account of auditory change detection. *Journal of Experimental Psychology: Human Perception and Performance*, 36, 994–1004.
- McDonald, K. L., & Alain, C. (2005). Contribution of harmonicity and location to auditory object formation in free field: Evidence from event-related brain potentials. *Journal of the Acoustical Society of America*, 118, 1593–1604.
- McFarland, D. J., & Cacace, A. T. (1995). Comparisons of memory for nonverbal auditory and visual sequential stimuli. *Psychological Research*, 57, 80–87.
- McKeown, D., & Wellsted, D. (2009). Auditory memory for timbre. *Journal of Experimental Psychology: Human Perception and Performance*, 35, 855–875.
- McLachlan, N., & Wilson, S. (2010). The central role of recognition in auditory perception: A neurobiological model. *Psychological Review*, 117, 175–196.
- Melara, R. D., & Marks, L. E. (1990). Interaction among auditory dimensions: Timbre, pitch, and loudness. *Perception & Psychophysics*, 48, 169–178.
- Mercer, T., & McKeown, D. (2010). Updating and feature overwriting in short-term memory for timbre. *Attention, Perception, & Psychophysics*, 72, 2289–2303.
- Merzenich, M. M., & Brugge, J. F. (1973). Representation of the cochlear partition of the superior temporal plane of the macaque monkey. *Brain Research*, 50, 275–296.
- Micheyl, C., Carlyon, R. P., Gutschalk, A., Melcher, J. R., Oxenham, A. J., Rauschecker, J. P., et al. (2007). The role of auditory cortex in the formation of auditory streams. *Hearing Research*, 229, 116–131.
- Mitroff, S. R., Simons, D. J., & Levin, D. T. (2004). Nothing compares 2 views: Change blindness can occur despite preserved access to the changed information. *Perception & Psychophysics*, 66, 1268–1281.
- Mondor, T. A., Zatorre, R. J., & Terrio, N. A. (1998). Constraints on the selection of auditory information. *Journal of Experimental Psychology: Human Perception and Performance*, 24, 66–79.
- Moore, B. C. J., & Gockel, H. (2002). Factors influencing sequential stream segregation. *Acta Acustica United with Acustica*, 88, 320–333.

- Näätänen, R., & Winkler, I. (1999). The concept of auditory stimulus representation in cognitive neuroscience. *Psychological Bulletin*, 125, 826–859.
- Nairne, J. S. (2003). Sensory and working memory. In A. F. Healy & R. W. Proctor (Eds.), *Handbook of psychology* (Vol. 4, pp. 423–444). Hoboken, NJ: Wiley.
- Nelken, I. (2004). Processing of complex stimuli and natural scenes in the auditory cortex. *Current Opinion in Neurobiology*, 14, 474–480.
- Oliva, A., & Torralba, A. (2001). Modeling the shape of the scene: A holistic representation of the spatial envelope. *International Journal of Computer Vision*, 42, 145–175.
- O'Regan, J. K., Deubel, H., Clark, J. J., & Rensink, R. (2000). Picture changes during blinks: Looking without seeing and seeing without looking. *Visual Cognition*, 7, 191–211.
- O'Regan, J. K., Rensink, R., & Clark, J. J. (1999). Change blindness as a result of 'mudsplashes'. *Nature*, 398, 34.
- Paivio, A., & Csapo, K. (1973). Picture superiority in free recall: Imagery or dual coding? *Cognitive Psychology*, 5, 176–206.
- Palmeri, T. J., & Gauthier, I. (2004). Visual object understanding. *Nature Reviews Neuroscience*, 5, 291–303.
- Pavani, F., & Turatto, M. (2008). Change perception in complex auditory scenes. *Perception & Psychophysics*, 70, 619–629.
- Penney, C. G. (1975). Modality effects in short-term memory. *Psychological Bulletin*, 82, 68–84.
- Penney, C. G. (1989). Modality effects and the structure of short-term verbal memory. *Memory & Cognition*, 17, 398–422.
- Phillips, W. A. (1974). On the distinction between sensory storage and short-term visual memory. *Perception & Psychophysics*, 16, 283–290.
- Polich, J. (2007). Updating P300: An integrative theory of P3a and P3b. *Clinical Neurophysiology*, 118, 2128–2148.
- Pourtois, G., De Preto, M., Hauert, C.-A., & Vuilleumier, P. (2006). Time course of brain activity during change blindness and change awareness: Performance is predicted by neural events before change onset. *Journal of Cognitive Neuroscience*, 18, 2108–2129.
- Pratt, H., Michalewski, H. J., Patterson, J. V., & Starr, A. (1989). Brain potentials in a memory-scanning task: II. Effects of aging on potentials to the probes. *Electroencephalography and Clinical Neurophysiology*, 72, 507–517.
- Protzner, A. B., Cortese, F., Alain, C., & McIntosh, A. R. (2009). The temporal interaction of modality specific and process specific neural networks supporting simple working memory tasks. *Neuropsychologia*, 47, 1954–1963.
- Pulvermuller, F. (2005). Brain mechanisms linking language and action. *Nature Reviews Neuroscience*, 6, 576–582.
- Rensink, R. A., O'Regan, J. K., & Clark, J. J. (1997). To see or not to see: The need for attention to perceive changes in scenes. *Psychological Science*, 8, 368–373.
- Roberts, L. A. (1986). Modality and suffix effects in memory for melodic and harmonic musical materials. *Cognitive Psychology*, 18, 123–157.
- Rugg, M. D., & Curran, T. (2007). Event-related potentials and recognition memory. *Trends in Cognitive Sciences*, 11, 251–257.
- Ryan, J. D., & Cohen, N. J. (2004). The nature of change detection and online representation of scenes. *Journal of Experimental Psychology: Human Perception and Performance*, 30, 988–1015.
- Samuel, A. G., & Kraljic, T. (2009). Perceptual learning for speech. *Attention, Perception, & Psychophysics*, 71, 1207–1218.
- Schankin, A., & Wascher, E. (2008). Unvoluntary attentional capture in change blindness. *Psychophysiology*, 45, 742–750.
- Schendel, Z. A., & Palmer, C. (2007). Suppression effects on musical and verbal memory. *Memory & Cognition*, 35, 640–650.
- 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.
- Semal, C., & Demany, L. (1991). Dissociation of pitch from timbre in auditory short-term memory. *Journal of the Acoustical Society of America*, 89, 2404–2410.
- Shamma, S. (2001). On the role of space and time in auditory processing. *Trends in Cognitive Sciences*, 5, 340–348.
- Simons, D. J., & Levin, D. T. (1998). Failure to detect changes to people during a real-world interaction. *Psychonomic Bulletin & Review*, 5, 644–649.
- Sinnett, S., Costa, A., & Soto-Faraco, S. (2006). Manipulating inattention blindness within and across sensory modalities. *Quarterly Journal of Experimental Psychology*, 59, 1425–1442.
- Snyder, J. S., & Alain, C. (2007). Toward a neurophysiological theory of auditory stream segregation. *Psychological Bulletin*, 133, 780–799.
- Sperling, G. (1960). The information available in brief visual presentations. *Psychological Monographs: General and Applied*, 74, 1–29.
- Squire, L. R., Stark, C. E. L., & Clark, R. E. (2004). The medial temporal lobe. *Annual Review of Neuroscience*, 27, 279–306.
- Squire, L. R., & Zola-Morgan, S. (1988). Memory: Brain systems and behavior. *Trends in Neurosciences*, 11, 170–175.
- Standing, L. (1973). Learning 10,000 pictures. *Quarterly Journal of Experimental Psychology*, 25, 207–222.
- Stolz, J. A., & Jolicoeur, P. (2004). Changing features do not guide attention in change detection: Evidence from a spatial cuing paradigm. *Psychonomic Bulletin & Review*, 11, 870–875.
- Thompson, R. F., & Kim, J. J. (1996). Memory systems in the brain and localization of a memory. *Proceedings Of The National Academy Of Sciences*, 93, 13438–13444.
- Trainor, L. J., & Trehub, S. E. (1992). A comparison of infants' and adults' sensitivity to western musical structure. *Journal of Experimental Psychology: Human Perception and Performance*, 18, 394–402.
- Triesch, J., Ballard, D. H., Hayhoe, M. M., & Sullivan, B. T. (2003). What you see is what you need. *Journal of Vision*, 3, 86–94.
- Ullman, S. (2007). Object recognition and segmentation by a fragment-based hierarchy. *Trends in Cognitive Sciences*, 11, 58–64.
- Visscher, K. M., Kaplan, E., Kahana, M. J., & Sekuler, R. (2007). Auditory short-term memory behaves like visual short-term memory. *PloS Biology*, 5, 662–672.
- 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.
- Ward, G., Avons, S. E., & Melling, L. (2005). Serial position curves in short-term memory: Functional equivalence across modalities. *Memory*, 13, 308–317.
- Weinberger, N. M. (2004). Specific long-term memory traces in primary auditory cortex. *Nature Reviews Neuroscience*, 5, 279–290.
- Wolfe, J. M., Kluender, K. R., Levi, D. M., Bartoshuk, L. M., Herz, R. S., Klatzky, R. L., et al. (2008). *Sensation and perception* (2nd ed.). Sunderland, MA: Sinauer.
- Wright, B. A., & Zhang, Y. (2009). A review of the generalization of auditory learning. *Philosophical Transactions of the Royal Society B*, 364, 301–311.
- Zelinsky, G. J. (2003). Detecting changes between real-world objects using spatiochromatic filters. *Psychonomic Bulletin & Review*, 10, 533–555.