frontiers in
PSYCHOLOGY

REVIEW ARTICLE
published: 07 February 2012
doi: 10.3389/fpsyg.2012.00015

Attention, awareness, and the perception of auditory scenes

Joel S. Snyder[1]*, Melissa K. Gregg[1], David M. Weintraub[1] and Claude Alain[2]

[1] Department of Psychology, University of Nevada Las Vegas, Las Vegas, NV, USA
[2] The Rotman Research Institute, Baycrest Centre for Geriatric Care, Toronto, ON, Canada

Edited by:
Naotsugu Tsuchiya, RIKEN, Japan

Reviewed by:
Alexander Gutschalk, Universität Heidelberg, Germany
Hirohito M. Kondo, Nippon Telegraph and Telephone Corporation, Japan

***Correspondence:**
Joel S. Snyder, 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

Auditory perception and cognition entails both low-level and high-level processes, which are likely to interact with each other to create our rich conscious experience of soundscapes. Recent research that we review has revealed numerous influences of high-level factors, such as attention, intention, and prior experience, on conscious auditory perception. And recently, studies have shown that auditory scene analysis tasks can exhibit multistability in a manner very similar to ambiguous visual stimuli, presenting a unique opportunity to study neural correlates of auditory awareness and the extent to which mechanisms of perception are shared across sensory modalities. Research has also led to a growing number of techniques through which auditory perception can be manipulated and even completely suppressed. Such findings have important consequences for our understanding of the mechanisms of perception and also should allow scientists to precisely distinguish the influences of different higher-level influences.

Keywords: auditory scene analysis, multistability, change deafness, informational masking, priming, attentional blink

INTRODUCTION

Understanding conscious experience of the external world has been a pursuit of theorists since the early days of experimental psychology. For example, Wundt and Titchener were among those who used introspection of their own perceptions to try and arrive at the fundamental units of experience (Boring, 1953; Danzinger, 1980). However, since then perception science and other areas of experimental psychology and neuroscience have been dominated by more objective psychophysical methods of understanding perception that have as a consequence, or by design, pushed the inquiry of subjective experience to the background. This objective measurement of perception has provided exquisite information about our perceptual skills to detect, discriminate, and categorize particular stimuli, and the underlying neuro-computational mechanisms of these abilities.

Recently, however, theorists have made an important contribution to reviving the scientific study of consciousness, perhaps most notably by defining accessible empirical problems such as how to explain the generation of perceptual *awareness* or *consciousness* (Crick and Koch, 1995, 2003), which we operationally define as the explicit reporting of a particular stimulus or how it is perceptually organized. This has led to investigations into the necessary and sufficient conditions for people to be aware of stimuli, especially in visual perception. For example, researchers have investigated the role of particular brain areas (Leopold and Logothetis, 1999; Tong et al., 2006; Donner et al., 2008) and particular neural processes such as feedback from higher to lower areas (Pascual-Leone and Walsh, 2001; Hochstein and Ahissar, 2002; Lamme, 2004; Wibral et al., 2009) that are associated with visual awareness. In many cases, these investigations have made use of multistable visual stimuli that can be perceived in more than one way (e.g., the well-known Necker cube, Long and Toppino, 2004), enabling the investigation of changes in perception without any confounding stimulus changes. The development of techniques to manipulate whether people are aware of particular stimuli (Kim and Blake, 2005) has additionally led to evaluating awareness (e.g., of a prior stimulus) as an independent variable (i.e., rather than studying awareness as the outcome variable) that can affect perception of subsequent stimuli (e.g., Kanai et al., 2006; for a review Koch and Tsuchiya, 2007). Much less work of these types has been done on auditory awareness, but several promising lines of research have begun, which we discuss in detail below.

In this review of the literature, we focus on three main types of research on auditory perception. First, we review research that demonstrates effects of attention and other high-level factors on auditory perceptual organization, with an emphasis on the difficulty in manipulating attention separately from other factors. Next, we discuss the fact that perception of sound objects exhibits the hallmarks of multistability and therefore shows promise for future studies of auditory perception and its underlying neural mechanisms. In this section, we also review research on the neural correlates of subjective auditory perception, which provides clues as to the areas of the brain that determine perception of sound objects. Finally, we discuss a number of recent demonstrations in which auditory events can be made imperceptible, which like their visual counterparts can enable researchers to identify the mechanisms of auditory awareness. Some of the studies that have been done permit interesting comparisons between perception of sound and conscious perception of stimuli in other sensory modalities. When possible, we will point out the similarities and differences across modalities, and point out the need for future research to delineate the extent to which similar phenomena

and similar mechanisms are present across the senses during perception.

AUDITORY SCENE ANALYSIS AS A FRAMEWORK TO STUDY AWARENESS

Auditory scene analysis (ASA) is a field of study that has been traditionally concerned with how the auditory system perceptually organizes incoming sounds from different sources in the environment into sound objects or streams, such as discrete sounds (e.g., phone ringing, gunshot) or sequences of sounds (e.g., melody, voice of a friend in crowded restaurant), respectively (Bregman, 1990). For example, in a crowded restaurant in which many people are talking at the same time, an individual must *segregate* the background speech from his or her dining partner's speech and *group* the various sound components of the partner's speech appropriately into a meaningful stream of words. ASA has mainly been studied with the goal of understanding how listeners segregate and group sounds; however, research in this field has also developed paradigms that are highly suitable for studying more general perceptual mechanisms and how low-level stimulus factors and higher-level factors such as attention, intention, and previous knowledge influence perception. In ASA studies, participants are often asked to report on their subjective experience of hearing two or more segregated patterns; and as mentioned earlier, when sounds are kept constant the operation of perceptual mechanisms can be studied directly without confounding effects of stimulus manipulations. However, indirect performance-based measures of segregation can also be informative because they tend to show the same effects as subjective measures (e.g., Roberts et al., 2002; Stainsby et al., 2004; Micheyl and Oxenham, 2010). Another important aspect of many ASA studies is that they often use rather simple arrangements of sounds that are easy to generate and manipulate. They also do not involve many of the complications associated with using real-world sounds (e.g., speech and music), such as the activation of long-term memory or expertise-related processes. Thus, such high-level processes can be controlled and studied with relative ease.

Bregman (1990) proposed two main classes of ASA mechanisms: (1) *primary* mechanisms that process incoming mixtures of sounds in an automatic fashion using simple transformations, and (2) *schema-based* mechanisms that are more likely to be attention-, intention-, and knowledge-dependent. An example of the operation of primary ASA is the well-known effect of frequency separation (Δf) during segregation of sequential tone patterns (Miller and Heise, 1950; Bregman and Campbell, 1971; Van Noorden, 1975). In the laboratory, auditory stream segregation has been studied extensively as an example of sequential segregation by playing two alternating pure tones of different frequencies (A and B) in a repeating pattern (e.g., ABA-ABA-. . ., where "-" corresponds to a silence), as shown in **Figure 1**. At first, the tones are heard as a single stream with a galloping rhythm, but after several repetitions of the sequence, the tones are often heard as splitting into two streams or "streaming" (i.e., A-A-A-A. . . and B—B—. . .). The larger the Δf between the A and B tones and the more rapidly they are presented, the more likely participants report hearing two streams as opposed to one stream. The characteristic time course of pure-tone streaming, called *buildup*, is likely to have its basis in the adaptation of frequency-tuned neurons in early brainstem and/or primary cortical stages of processing (Micheyl et al., 2005; Pressnitzer et al., 2008; for reviews, Micheyl et al., 2007a; Snyder and Alain, 2007). But more recent research has shown that a number of stimulus cues besides pure-tone frequency can result in perception of streaming, even cues that are known to be computed in the central auditory system (for reviews, Moore and Gockel, 2002; Snyder and Alain, 2007). This evidence that streaming occurs at central sites raises the possibility that auditory perception results from a combination of activity at multiple levels of the auditory system, including those that can be influenced by schema-based mechanisms.

In addition to segregation of sequential patterns, another important aspect of scene analysis is the segregation of sounds

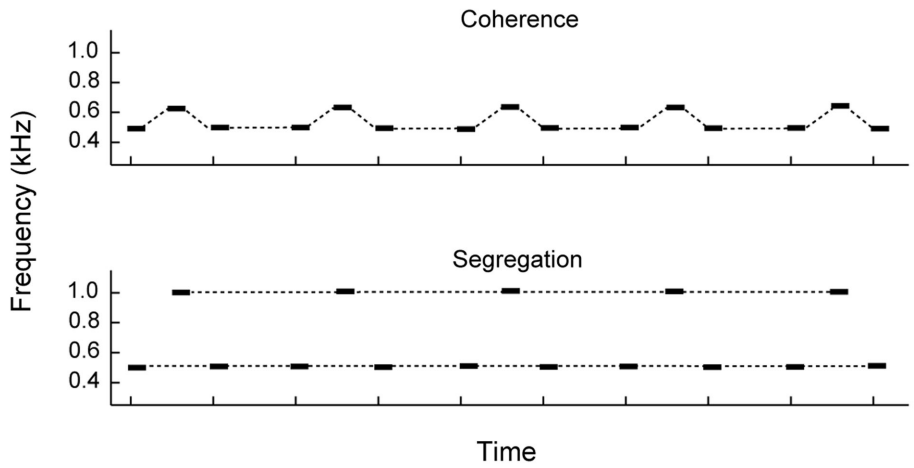

FIGURE 1 | In auditory stream segregation experiments, low and high tones are alternated repeatedly. When the frequency difference between the tones is small (top), this typically leads to perception of one coherent stream. For large frequency differences (bottom), one is more likely to be heard as two segregated streams.

that occur concurrently, such as when two individuals speak at exactly the same time. In social gatherings, human listeners must perceptually integrate the simultaneous components originating from one person's voice (i.e., fundamental frequency or f_0, and harmonics that are integer-multiples of f_0) and segregate these from concurrent sounds of other talkers. Psychophysical research has identified several cues that influence how concurrent sounds will be grouped together (for reviews, Carlyon, 2004; Alain, 2007; Ciocca, 2008). For instance, sounds that are harmonically related, begin at the same time and originate from the same location are more likely to emanate from the same physical object than those that are not. In the laboratory, experimenters can induce the perception of concurrent sound objects by mistuning one spectral component (i.e., a harmonic) from an otherwise periodic harmonic complex tone (see **Figure 2**). Low harmonics mistuned by about 4–6% of their original value stand out from the complex so that listeners report hearing two sounds: a complex tone and another sound with a pure-tone quality (Moore et al., 1986). While several studies have investigated the role that attention plays in auditory stream segregation, which we review below, far less research has been done on the impact of high-level factors on concurrent sound segregation.

EFFECTS OF HIGH-LEVEL FACTORS ON AUDITORY SCENE ANALYSIS

Attention

Attention during auditory stream segregation. Psychophysical studies have shown that buildup of stream segregation is modulated by attention, suggesting the involvement of high-level factors in perception of streaming. In these studies, participants were presented with an ABA- pattern to one ear. The role attention plays in auditory stream segregation was examined by assessing the buildup of streaming while participants were engaged in a separate auditory, visual, or non-sensory task in which participants counted backward (Carlyon et al., 2001, 2003; Thompson et al., 2011). By having participants engaging a primary task, attention was diverted away from the ABA- pattern. When attending to the ABA- pattern, participants showed a typical pattern of buildup. However, when attending the other task for the first part of the ABA- pattern, participants failed to show any sign of buildup when they switched their attention. Thus, buildup either did not occur while attention was diverted to the primary task or it was reset following the brief switch in attention (Cusack et al., 2004), a distinction that has been quite difficult to resolve using psychophysical measurements. These effects occurred regardless of the task used to capture attention (Carlyon et al., 2003), suggesting that buildup involves mechanisms within central auditory areas, multimodal pathways, and/or in peripheral areas that can be influenced in a top-down fashion by attention. To explain these results, Cusack et al. (2004) proposed a hierarchical model of stream segregation. According to this model, preattentive mechanisms segregate streams based on acoustic features (e.g., Δf) and attention-dependent buildup mechanisms further break down outputs (streams) of this earlier process that are attended to. For example, when talking to a friend at a concert, low-level processes automatically segregate the friend's voice from the music. However, since attention is allocated to the friend's voice and not the concert, buildup processes do not further decompose the music into its constituent parts (e.g., guitar, drums, bass, etc.; also, see Alain and Arnott, 2000).

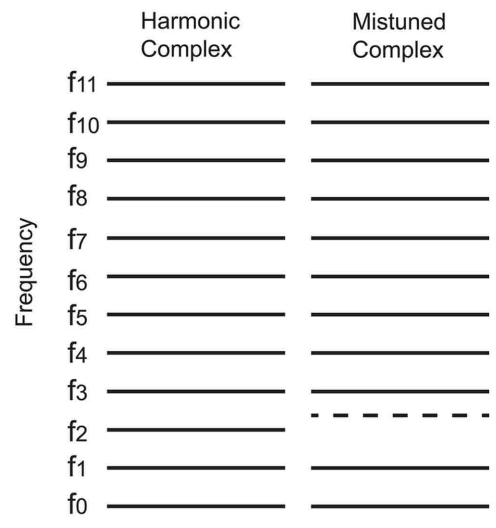

FIGURE 2 | In mistuned harmonic experiments, a complex harmonic sound composed of frequency components that are all multiples of the fundamental frequency (f_0) is heard as a single sound with a buzzy quality (left). When one of the components is mistuned, it stands out as a separate pure-tone object in addition to the remaining complex sound (right).

Consistent with this model, Snyder et al. (2006) provided event-related potential (ERP) evidence for at least two mechanisms contributing to stream segregation: an early preattentive segregation mechanism and an attention-dependent buildup mechanism. In particular, auditory cortical responses (P2 and N1c) to an ABA- pattern increased in amplitude with increasing Δf and correlated with behavioral measures of streaming; this enhancement occurred even when attention was directed away from the ABA- pattern. Additionally, a temporally broad enhancement following the onset of an ABA- pattern progressively increased in positivity throughout the course of the pattern. The time course of this progressive increase indicated a strong link with the buildup of streaming. Importantly, this enhancement was diminished when participant's attention was directed away from the ABA- pattern. These findings support the existence of an attention-dependent buildup mechanism in addition to a preattentive segregation mechanism. Also, since buildup-related processes were measured during passive listening these findings are more consistent with an effect of sustained attention as opposed to the possibility that buildup is simply reset following brief switches in attention (cf. Cusack et al., 2004).

However, Sussman et al. (2007) showed that buildup does not always require attention. They showed that deviant stimuli embedded within a high-tone stream of an ABA- pattern resulted in a mismatch negativity response during perception of two streams (Sussman et al., 1999, 2007). Furthermore, deviants were more likely to evoke a mismatch negativity when they occurred at the end of ABA- patterns compared to when they occurred early on, consistent with the time course of buildup. Importantly, these findings were similar whether or not the ABA- patterns were attended,

suggesting that attention may not be required for buildup to occur, in contrast to the findings discussed above. Because this study used relatively large Δfs, it is possible that attention only modulates buildup in the absence of robust segregation cues (i.e., large Δf; Sussman et al., 2007). Indeed, Snyder et al. (2006) included several conditions with Δfs smaller than that used by Sussman et al. (2007). Additionally, close inspection of Cusack et al. (2004) shows that preattentive buildup processes were more prevalent for larger than smaller Δf conditions.

Several additional physiological studies have examined the effects of selective attention on streaming. These studies have supported a *gain* model in which attention to a target stream enhances neural processing of sounds within that stream while suppressing unattended streams. An early ERP study showed that selective attention to a stream facilitated early sensory processing of that stream and inhibited processing of unattended streams (Alain and Woods, 1994). More recent studies have focused on the effects of selective attention on continuous neural activity to sound streams. For example, in addition to enhanced transient responses generated in associative auditory areas (Bidet-Caulet et al., 2007), selective attention enhanced steady-state responses generated in primary auditory cortex to attended streams (Bidet-Caulet et al., 2007; Elhilali et al., 2009b; Xiang et al., 2010). Furthermore, these responses were entrained to the rhythm of the target stream and constrained by known entrainment capabilities within auditory cortex (i.e., better entrained for low vs. high frequencies; Xiang et al., 2010). High-density ERP and neuromagnetic studies have recently examined neural responses to continuous speech streams played amongst distracting speech (Kerlin et al., 2010; Ding and Simon, 2012). Both studies demonstrated that low-frequency (4–8 Hz) speech envelope information was represented in the auditory cortex of listeners. These representations were measured as either a continuous low-frequency response phase-locked to the speech (Kerlin et al., 2010) or a phase-locked N1-like neuromagnetic response that was primarily driven by low-frequency features of the speech (Ding and Simon, 2012). Consistent with a gain model, selectively attending to a speech stream enhanced the continuous low-frequency response to the attended speech and (possibly) suppressed responses to unattended speech (Kerlin et al., 2010). In a separate study, attention enhanced an N1-like response to attended speech and suppressed responses to unattended speech (Ding and Simon, 2012). In this latter case, the relatively short latency of these effects suggests that attention modulated bottom-up segregation and/or selection processes. Furthermore, this finding generalizes similar effects of selective attention on the auditory N1 ERP response from simple tones (Hillyard et al., 1973) to more naturalistic speech stimuli. Taken together, these findings are consistent with a gain model in which attention to a sound stream improves its neural representation while suppressing representations of irrelevant streams.

An issue with this type of gain model is that it is not uncommon for separate streams of speech to share similar acoustic features and, accordingly, activate overlapping neuronal receptive fields. In this case, attention-related enhancement or suppression would act on both attended and unattended streams. Therefore, in addition to gain, attention may also serve to narrow neuronal receptive fields of neurons within the auditory cortex (Ahveninen et al., 2011). This would, in effect, increase feature selectivity and decrease the likelihood that separate streams of speech activate overlapping neurons. To test this model, participants were presented with target sequences of repeating tones embedded within notch-filtered white noise that did not overlap with the frequency of the target. Auditory cortical responses (N1) to unattended sounds were reduced in amplitude reflecting lateral inhibition from the masker. In contrast, these attenuated effects disappeared for attended target stimuli. Here, selective attention may have narrowed the width of the receptive fields processing the target stream and, consequently, increased the representational distance between task-relevant and task-irrelevant stimuli. Furthermore, these neuronal changes correlated with behavioral measures of target detection suggesting that attention-related receptive field narrowing aided segregation, in addition to any helpful effects of gain.

A third way in which selective attention influences neural processes of streaming is enhancing temporal coherence between neuronal populations. In particular, attention to a target stream enhanced synchronization between distinct neuronal populations (both within and across hemispheres) responsible for processing stimuli within that stream and this correlated with behavioral measures of streaming (Elhilali et al., 2009b; Xiang et al., 2010). Enhanced synchronization may have facilitated the perceptual boundary between acoustic features belonging to attended and unattended streams as detected by a temporal coherence mechanism (Shamma et al., 2011). Consistent with the role of temporal coherence in streaming, when presented with a modified ABA- pattern in which low- (A) and high- (B) pitched tones were played simultaneously rather than sequentially participants reported hearing one stream even for very large Δfs (Elhilali et al., 2009a). Taken together, these physiological studies revealed at least three ways in which attention modulated streaming: (1) enhanced processing of the stimuli within the task-relevant stream and suppressed processing of those within the task-irrelevant stream (Alain and Woods, 1994; Bidet-Caulet et al., 2007; Elhilali et al., 2009b; Kerlin et al., 2010; Xiang et al., 2010; Ding and Simon, 2012), (2) enhanced feature selectivity for task-relevant stimuli (Ahveninen et al., 2011), and (3) enhanced temporal coherence between distinct neuron populations processing task-relevant stimuli (Elhilali et al., 2009b; Xiang et al., 2010).

Jones et al. (1981) theorized that rhythmic attention plays a role in the stream segregation process. Rhythmic attention is assumed to be a time-dependent process that dynamically fluctuates in a periodic fashion between a high and low state (Large and Jones, 1999). According to this theory, rhythmic attention aids listeners in picking up relations between adjacent and non-adjacent events when they are nested in a common rhythm. Therefore, when stimuli have a regular periodic pattern, rhythmic attention can detect sounds that do and do not belong to that stream. Indeed, when two streams of tones differed in rhythm they were more likely to be segregated even for tones relatively close in frequency (Jones et al., 1981). These findings are consistent with physiological studies that showed steady-state brain responses to be entrained to the rhythm of the segregated target stream (Elhilali et al., 2009b; Xiang et al., 2010). However, follow-up studies to Jones et al. (1981) have yielded conflicting results. For example, Rogers and Bregman (1993) showed that the likelihood of a context

sequence of B-only tones to increase segregation in a short ABA- pattern was similar for context sequences that matched or mismatched the ABA- rhythm. Therefore, manipulating rhythm only minimally enhanced the effect of Δf during perception of streaming. However, it is not clear whether the buildup observed during these single-tone contexts was mediated by similar mechanisms as those that are active while listening to an ABA- context pattern (Thompson et al., 2011). Therefore, it is possible that rhythmic attention modulates these two types of buildup in a different manner. Studies by Alain and Woods (1993, 1994) also provided little evidence that rhythm has a role in streaming. They showed that the likelihood of segregating a target stream of tones from distracters was similar for sequences that had regular or irregular rhythms. However, because the rhythms of target and distracter streams were never manipulated independently, rhythm could not be used as a reliable cue for segregation. Therefore, in light of these issues, it still seems possible that rhythmic attention may modulate stream segregation, especially in cases where Δf is not sufficient for segregation to occur.

Indeed, the role of rhythmic attention in streaming has been the focus of several recent studies, which have proven more consistent with the ideas of Jones et al. (1981). For example, when Δfs were small, listeners were more likely to segregate an irregular target stream from a distracter stream when the distracter was isochronous (Andreou et al., 2011). However, given a large enough Δf, rhythm had a marginal influence on measures of streaming. Therefore, it may be that large Δfs are a dominant cue for streaming, but that listeners consider other cues such as rhythm when Δf is small. Other studies, in which participants detected a target melody interleaved with irrelevant melodies, showed that participants used rhythmic pattern to attend to points in time during which notes of the target melody occurred (Dowling et al., 1987) and reduce the distracting effects of irrelevant melodies (Devergie et al., 2010). Finally, listeners used rhythmic differences between streams to maintain perception of segregated streams (Bendixen et al., 2010). A plausible explanation for these results is that attention to the task-relevant stream was facilitated when the target stream had a regular rhythm distinct from other streams. Additionally, increased suppression of isochronous distracter streams facilitated attention to an irregular task-relevant stream. Taken together, studies suggest that rhythmic attention may modulate streaming, perhaps in conditions in which more salient cues are unavailable, but more work is needed to assess the generality of these findings.

Attention during concurrent sound segregation. As with stream segregation, scalp-recorded ERPs have proven helpful in investigating the role of attention during concurrent sound perception because it allows one to examine the processing of auditory stimuli while they occur outside the focus of attention. Alain et al. (2001) measured auditory ERPs while participants were presented with harmonic complex tones with or without a mistuned harmonic; in one condition they indicated whether they heard one vs. two sounds, while in another condition they listened passively (i.e., read a book of their choice, with no response required). The main finding was an increased negativity that superimposed the N1 and P2 wave elicited by the sound onset. **Figure 3** shows examples of neuromagnetic activity elicited by tuned and mistuned stimuli and the corresponding difference wave referred to as the object-related negativity (ORN), so named because its amplitude correlated with the observers' likelihood of hearing two concurrent sound objects. The ERP recording by Alain et al. (2001) during the passive listening condition was instrumental in showing that the ORN, thought to index concurrent sound segregation and perception, occurred automatically. The proposal that low-level concurrent sound segregation mechanisms are not under attentional control was confirmed in subsequent ERP studies using active listening paradigms that varied auditory (Alain and Izenberg, 2003) or visual attentional demands (Dyson et al., 2005).

In addition to providing evidence for primary sound segregation, ERPs also revealed attention-related effects during the perception of concurrent sound objects. Indeed, when listeners were required to indicate whether they heard one or two sounds, the ORN was followed by a positive wave that peaked about 400 ms after sound onset, referred to as the P400 (Alain et al., 2001). It was present only when participants were required to make a response about the stimuli and hence is thought to index perceptual decision-making. Like the ORN, the P400 amplitude correlated with perception and was larger when participants were more likely to report hearing two concurrent sound objects. Together, these ERP studies revealed that both bottom-up (attention-independent) and top-down controlled processes are involved in concurrent sound perception.

In the ERP studies reviewed above, the perception of concurrent sound objects and mistuning were partly confounded, making it difficult to determine whether the ORN indexes conscious perception or simply the amount of mistuning. If the ORN indexes perception of concurrent sound objects, then it should also be present when concurrent sounds are segregated on the basis of other cues such as spatial location. McDonald and Alain (2005) examined the role of location on concurrent sound perception. Using complex harmonic tones with or without a mistuned harmonic, these authors found that the likelihood of reporting two concurrent sound objects increased when the harmonic was presented at a different location than the remaining harmonics of the complex. Interestingly, the effect of spatial location on perception of concurrent sound objects was paralleled by an ORN. The results from this study indicated that the ORN was not limited to mistuning but rather relates to the subjective experience of hearing two different sounds simultaneously. Moreover, this study showed that listeners can segregate sounds based on harmonicity or location alone and that a conjunction of harmonicity and location cues contributes to sound segregation primarily when harmonicity is ambiguous. Results from another research group also found an ORN during concurrent sound segregation with cues other than harmonicity, further supporting the interpretation that the ORN is related to conscious perception rather than stimulus processing (Johnson et al., 2003; Hautus and Johnson, 2005). However, an even stronger test of this account would be to present multistable versions of the mistuned harmonic (i.e., with an intermediate amount of mistuning) to see if the ORN is enhanced when listeners hear two objects compared to when they hear one object for the exact same stimulus.

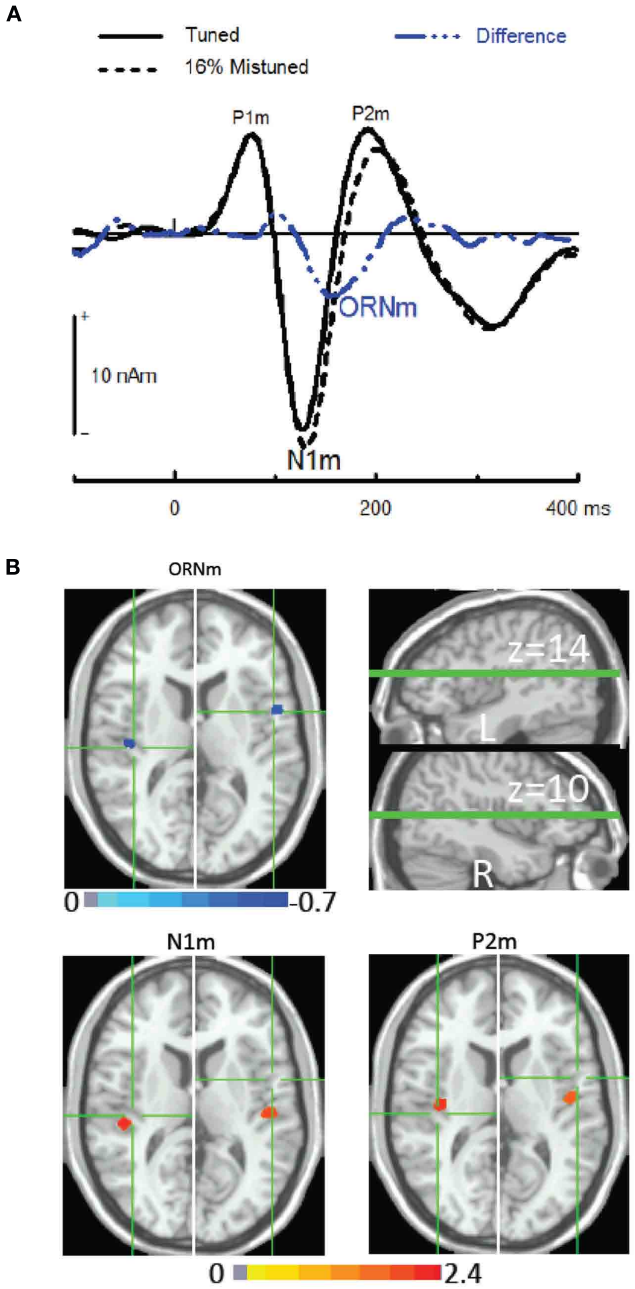

FIGURE 3 | A neural marker of concurrent sound segregation based on harmonicity. (A) Neuromagnetic activity elicited by harmonic complexes that had all harmonics in tune or the third harmonic mistuned by 16% of its original value. The magnetic version of the ORN (ORNm) is isolated in the difference wave between responses elicited by the tuned and mistuned stimuli. The group mean responses are from 12 young adults. **(B)** Source modeling using the beamforming technique called event-related synthetic aperture magnetometry (ER-SAM). The activation maps (group image results) are overlaid on a brain image conforming to Talairach space. Green cross hairs highlight the location of the peak maxima for the ORNm sources (blue) derived from subtracting the ER-SAM results for the 0% mistuned stimulus from that of the 16% mistuned stimulus. For comparison, ER-SAM source maps at the time interval of the peak N1m and P2m responses (red) are plotted at the same ORNm axial (*z*-plane) level.

Though the mistuned harmonic paradigm has proven helpful in identifying neural correlates of concurrent sound perception, the conclusions from these studies often rely on subjective assessment. Moreover, it is unclear whether the mechanisms involved in parsing a mistuned harmonic in an otherwise harmonic complex share similarities with those involved during the segregation and identification of over-learned stimuli such as speech sounds. In addition to data-driven processes, speech stimuli are likely to engage schema-driven processes during concurrent speech segregation and identification. To examine whether prior findings using the mistuned harmonic paradigms were generalizable to more ecologically valid stimuli, Alain et al. (2005) recorded ERPs while participants performed the double vowel task. The benefit of this task is that it provides a more direct assessment of speech separation and also evokes processes involved in acoustic identification. Here, participants were presented with a mixture of two phonetically different synthetic vowels, either with the same or different f_0, and participants were required to indicate which two vowels were presented. As previously reported in the behavioral literature (e.g., Chalikia and Bregman, 1989; Assmann and Summerfield, 1990), accuracy in identifying both vowels improved by increasing the difference in the f_0 between the two vowels. This improvement in performance was paralleled by an ORN that reflected the difference in f_0 between the two vowels. As with the mistuned stimuli, the ORN during speech segregation was present in both attend and ignore conditions, consistent with the proposal that concurrent speech segregation may involve an attention-independent process. In summary, while it is not yet possible to propose a comprehensive account of how the nervous system accomplishes concurrent sound segregation, such an account will likely include multiple neuro-computational principles and multiple levels of processing in the central auditory system.

Intention

One of the first in-depth investigations of streaming provided an elegant demonstration of the large influence that manipulating an observer's intention can have on conscious perception (Van Noorden, 1975). Participants listened to an ABA- pattern in which the A tone started out being much higher (or lower) than the B tone and increased (or decreased) in frequency after each presentation while the B tone stayed constant. This resulted in a continuously changing Δf between the A and B tones, and thus a continuously changing likelihood of hearing one or two streams. The stimulus onset asynchrony (SOA) from one tone to the next was also varied to promote auditory streaming. In addition to the changing stimulus, the participants' intention varied as a result of the following instructions: (1) try and hear a single stream, or (2) try and hear two streams. The participants' continuous "one stream" vs. "two streams" responses as a function of Δf and SOA provided a way to assess the limits of hearing a sequence of tones as integrated or segregated (see **Figure 4**). The Δf at which it was no longer possible to hold this percept was called the "fission boundary," and did not vary much with SOA. In contrast, when participants were asked to hold the one stream percept, the Δf at which it was no longer possible (the "temporal coherence boundary") varied substantially with SOA. Importantly, these two perceptual boundaries did not overlap with each other, resulting in a large number of

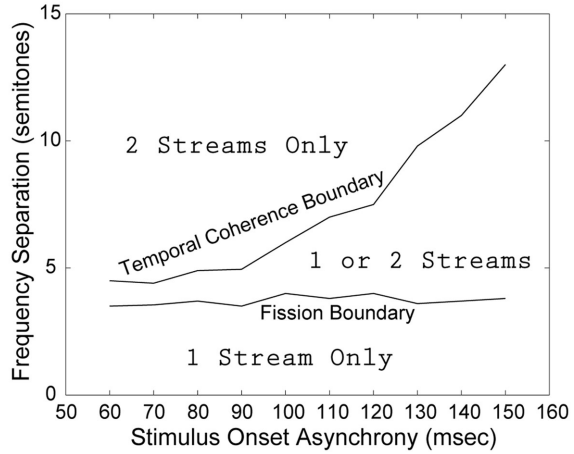

FIGURE 4 | A reproduction of Van Noorden's (1975) streaming diagram, showing the combinations of frequency separation and stimulus onset asynchrony between low and high tones that lead to perception of only one stream, only two streams, or either perceptual organization.

combinations of Δf and SOA in which either percept was possible. Not only did this demonstrate the large effect intention can have on conscious perception, it also was suggestive of other properties associated with conscious visual perception such as hysteresis and multistability (cf. Hock et al., 1993), foreshadowing more recent research to be discussed in detail below. Interestingly, Van Noorden used the term "attentional set" instead of "intention" to describe the manipulated variable in his study, which raises the important possibility that the effects he observed were due most directly to the scope of selective attention of the listener on either both the A and B tones or just one of the tones. Thus, while selective attention may be a mediating mechanism for the effect of intention to hear a particular perceptual organization on perception, it might not be the only way that a listener's intention can affect conscious perception. Given that, surprisingly little research has been done since Van Noorden's study to distinguish between effects of attention and intention, at either the behavioral or neurophysiological level, this remains a rich area to be investigated further.

Prior experience

One way to study higher-order cognitive processes during perception is to assess the impact of prior experience, which can inform the role of explicit and implicit memory during perception. For example, streaming studies have tested for effects of prior knowledge of stimuli as a possible mediating mechanism for a listener's intention to hear segregated patterns in an auditory scene. In one early study, listeners were presented with two melodies at the same time, with the tones of melody A interleaved with the tones of melody B (i.e., A_1, B_1, A_2, B_2,. . ., where A_1 is the first note of melody A). This results in a melody that is more complex than the typical ABA- pattern used for streaming experiments because the A and B tones frequently change during a trial (Dowling, 1973). When both melodies were familiar tunes, it was easier to identify them when the frequency ranges of the two melodies were greatly separated, as in standard streaming paradigms. Importantly, when the name of one of the tunes was given prior to hearing the interleaved melodies, it was easier to perceptually segregate it even when the two melodies were closer in pitch, demonstrating an effect of prior knowledge on perceptual segregation. However, knowing the name of the background melody did not help participants identify the target melody, suggesting that prior knowledge does not attenuate the distracting influence of background sounds (also, see Newman and Evers, 2007). Instead, it seems more likely that attentional focus upon expected notes in the target melody helped segregate it from the background. A later study directly tested this idea, showing that target melodies with events presented at points of high temporal expectation due to the rhythm of the A and B melodies were recognized better than melodies with events presented at points of low expectation (Dowling et al., 1987). This form of temporal attention is consistent with the dynamic attending theory of Jones and colleagues (Jones, 1976; Jones and Boltz, 1989; Large and Jones, 1999). A caveat to the work by Dowling on effects of familiarity is a more recent study showing that previously unfamiliar interleaved melodies were not easier to segregate when the target melody had just been presented by itself prior to the interleaved melodies (Bey and McAdams, 2002). Thus, the beneficial effects resulting from familiarity may only occur when the patterns are stored in long-term memory. Alternatively, it is possible that representations for familiar melodies are simply stronger than short-term traces for melodies that have been presented only once, regardless of the storage mechanism.

In some cases discussed thus far it is difficult to rule out attention as the most direct factor that enhances processing when manipulating familiarity of stimuli or the listener's intention. However, it is also possible that familiarity, priming, and other memory-related factors might be able to directly influence perception through non-attention-related mechanisms. For example, adults of all ages benefit from semantic predictability of words in a sentence segregation task (Pichora-Fuller et al., 1995). Another study showed that complex stimuli that are embedded in noise became easier to segregate when they were presented repeatedly, as long as on each new presentation they were mixed with a different noise (McDermott et al., 2011). Because the noises are unlikely to be perceived as auditory objects prior to the first time they are successfully segregated, this result suggests that short-term memory traces are able to automatically facilitate segregation. Finally, studies of streaming context effects have shown that both prior stimuli and prior perception of those stimuli can have large effects on a subsequent perceptual decision, an example of implicit memories influencing perception (Snyder et al., 2008, 2009a,b; Snyder and Weintraub, 2011; for similar findings in continuity perception, see Riecke et al., 2009, 2011; for related research in speech perception, see McClelland et al., 2006). In particular, a prior ABA- pattern with a large Δf biases following patterns to be heard as one stream, a *contrastive* or *suppressive* effect; in contrast, prior perception of two streams biases subsequent patterns to be heard with the same percept, a *facilitative* effect. Importantly, these streaming context effects are likely to be implicit because listeners are not explicitly asked to compare prior and current patterns nor are they typically aware that the prior patterns are affecting their perception. Also

of note is that the context effects are consistent with similar effects of prior stimuli and prior percepts observed in vision, suggesting the operation of general memory mechanisms that implicitly influence perception (Pearson and Brascamp, 2008).

One account of perception, known as reverse hierarchy theory, might help explain how high-level factors such as intention and prior experience might enhance segregation (Hochstein and Ahissar, 2002). This theory assumes that a stimulus activates the sensory system in a bottom-up manner without conscious access of each low-level representation; when the information finally reaches a high-level representation, this is accessed in the form of a gist or category related to the stimulus (also, see Oliva and Torralba, 2001; Greene and Oliva, 2009). Once this high-level activation occurs, low-level representations of the stimulus can be accessed only in specific circumstances in which a top-down path is possible. The reverse hierarchy theory is consistent with visual learning studies and a number of other visual phenomena including change blindness and illusory conjunctions (Hochstein and Ahissar, 2002). Recently, the theory was also able to predict novel findings in a word segregation task (Nahum et al., 2008). Hebrew-speaking listeners were unable to use a low-level binaural difference cue that would have aided segregation when the task was to make a semantic judgment on one of two possible Hebrew words that were phonologically similar (e.g., /tamid/ and /amid/). This was likely due to the fact that the semantic task primarily involved accessing high-level representations and the acoustically similar words were processed in highly overlapping ascending auditory pathways. Interestingly, even when the task was not inherently high-level such as in word identification, binaural cues were not used unless they were available on every trial within a block, suggesting that listeners implicitly learn over the course of a block of trials to not access low-level representations unless they were consistently useful. For our purposes, these results are interesting because they support a theory that might be able to explain how high-level information about stimuli and recent experience can guide the accessing of low-level cues for conscious auditory perception.

MULTISTABILITY

The fact that subjective and objective measures of perception can be substantially modulated by attention and other high-level factors suggests that auditory perception is multistable like visual perception (Leopold and Logothetis, 1999; Long and Toppino, 2004; Pearson and Brascamp, 2008). However, it was not until relatively recently that a thorough quantitative comparison was made between auditory and visual multistable perception (Pressnitzer and Hupé, 2006). In this study, the authors assessed auditory streaming using ABA- patterns with an intermediate Δf presented with many more repetitions per trial than usual. The same observers were also tested on perceptual segregation of moving plaid patterns, which has been studied in detail at the psychophysical (e.g., Hupé and Rubin, 2003) and neurophysiological (e.g., Movshon et al., 1985) level. Perception of the moving plaid pattern was appropriate for this comparison with perception of ABA- patterns because the two stimuli share a number of psychophysical properties. First, they are both segregation tasks, resulting in either the perception of a single pattern or two distinct patterns. Second, in both paradigms the initial perception is of a single pattern and only after a buildup period does perception of two patterns occur. The study went further by showing that after the initial switch to perceiving two patterns, observers then showed similar stochastic switching between the two percepts in both modalities. And the initial period of perceiving one stream was longer in duration than subsequent periods of either stable percept. They also showed that it was possible to intentionally control perception but it was not possible to completely eliminate switching between percepts, consistent with the findings of Van Noorden (1975) discussed earlier. The finding that even for ABA- patterns with rather large or rather small Δf values (i.e., not "ambiguous") switching between one and two streams continued to occur, despite an overall bias for one percept, emphasizes the robustness of multistability in streaming (Denham and Winkler, 2006).

Pressnitzer and Hupé (2006) further showed that despite the similar multistable perceptual phenomena in the visual and auditory paradigms, the number of switches per unit time in one modality did not predict the switching rate in the other modality, suggesting similar but functionally distinct mechanisms for controlling perception in vision and hearing. In a subsequent study, these authors further explored the mechanisms controlling multistable perception by presenting visual and auditory patterns at the same time (Hupé et al., 2008). In the first experiment, they presented ABA- and plaid patterns together and participants reported any switches observed in each modality. In the second experiment, they presented ABA- and apparent motion patterns together that were spatially and temporally coupled with each other, in order to increase the likelihood of cross-modal interactions in perception. The results showed that a switch in one modality did increase the likelihood of switching in the other modality, that the likelihood of perceiving the same percept in the two modalities was higher than expected based on chance, and these two effects were largest for the experiment using cross-modally coupled patterns. Thus, while there is likely to be interaction between the two modalities in controlling perception, this latter finding suggested that there is not a supramodal mechanism that controls perception in both modalities; rather, perceptual mechanisms in vision and hearing may interact depending on how likely signals in the two modalities are coming from the same physical objects in the environment. This conclusion is consistent with a study showing that intentional control over perceptual interpretations is strongly enhanced when stimuli are cross-modally consistent with each other (van Ee et al., 2009).

Neurophysiological studies also support the idea that perception may be determined primarily within modality-specific brain areas. In vision, the majority of findings show robust correlates in areas that are thought to be primarily dedicated to visual processing (Leopold and Logothetis, 1999; Tong et al., 2006). In hearing, although there are only a few studies on neural correlates of multistable perception, the findings also suggest the involvement of auditory-specific processes. However, it is important to be cautious in interpreting the precise role of brain areas measured in neurophysiological studies because of the correlational nature of the data.

In a streaming study measuring neuromagnetic brain activity signals from the superior temporal plane, small modulations in

sensory-evoked response amplitude were observed depending on whether listeners were hearing two streams. These perception-related modulations occurred in similar components as those that were modulated by increased Δf, but they were smaller in amplitude (Gutschalk et al., 2005). Intracranial ERPs from several lateral superior temporal lobe locations measured during neurosurgery in epilepsy patients also showed some dependence on perception, but these were also much less robust compared to Δf-dependent modulations (Dykstra et al., 2011). In a functional magnetic resonance imaging (fMRI) study, listeners showed more activity in auditory cortex when hearing two streams as opposed to one stream (Hill et al., 2011). In another fMRI study, which examined neural correlates of switching between one- and two-stream percepts, switching-related activations were observed in non-primary auditory cortex as well as the auditory thalamus in a manner that suggested the importance of thalamo-cortical interactions in determining perception (Kondo and Kashino, 2009). In an fMRI study on streaming using inter-aural time difference as the cue to segregating A and B tones, switching-related activity in the auditory cortex was again found, in addition to activity in the inferior colliculus, which is an important brainstem area for processing binaural information (Schadwinkel and Gutschalk, 2011). Future studies should directly compare the effect of perceiving one vs. two streams and the effect of switching between perceiving one and two streams; without such a direct comparison using the same participants and similar stimuli, it is difficult to determine whether similar brain circuits are implicated in these possibly distinct processes.

Exceptions to evidence for modality-specific auditory perception mechanisms are fMRI studies showing enhanced activity while perceiving two streams compared to perceiving one stream in the intraparietal sulcus, an area that is thought to also be involved in visual perceptual organization and attention shifting (Cusack, 2005; Hill et al., 2011). Interestingly, increasing the spectral coherence of complex acoustic stimuli in such a way that increases perceptual segregation also modulated the fMRI signals in intraparietal sulcus, in addition to the superior temporal sulcus, a higher-order auditory processing area (Teki et al., 2011). However, these brain modulations were observed while participants were not making perceptual judgments so it is unclear the extent to which they reflect perceptual processing, as opposed to automatic stimulus processing. At this point it is difficult to conclusively state which of the brain areas found to correlate with perception in these studies are most likely to be important for determining perception because of the different stimuli and tasks used. But these studies have provided a number of candidate areas that should be studied in future neurophysiological studies, as well as studies that assess the consequences of disrupted processing in the candidate areas.

Although other ASA tasks (e.g., mistuned harmonic segregation) have not been studied as thoroughly for signs of multistable perception observed in streaming, it stands to reason that they would show some of the same phenomena and could be useful in determining the generality of the streaming findings. For example, a multistable speech perception phenomenon is *verbal transformation* in which repeated presentation of a word results in the perceived word changing to another word, often with many different interpretations during a single trial (e.g., the four-phoneme stimulus TRESS being heard as the following sequence of words "stress, dress, stress, dress, Jewish, Joyce, dress, Jewess, Jewish, dress, floris, florist, Joyce, dress, stress, dress, purse"; Warren, 1968). Ditzinger and colleagues showed that rather than randomly changing between all the possible alternatives, pairs of alternatives tended to alternate with each other, suggesting that the principles underlying the phenomenon are more similar to other multistable phenomena (Ditzinger et al., 1997b; Tuller et al., 1997). Indeed, a dynamic systems model that was similar to a model of multistable visual perception was able to reproduce the time course of verbal transformations (Ditzinger et al., 1997a).

A more recent study took a different theoretical approach to verbal transformations by trying to explain them in terms of auditory streaming and grouping mechanisms (Pitt and Shoaf, 2002). Listeners were presented three-phoneme (consonant–vowel–consonant) pseudowords and reported instances of hearing transformations in addition to instances of hearing more than one stream of sounds. A large majority of the transformations reported were accompanied by hearing more than one stream of sounds, suggesting that part of the original pseudoword was segregated from the remainder, changing how the remainder sounded. Changes in perception also occurred for sine-wave speech that was repeated, with transformations occurring after more stimulus repetitions when perceived as speech rather than as tones, suggesting an influence of top-down knowledge on stabilizing perception, consistent with evidence from streaming paradigms discussed above. Behavioral evidence that overt and covert speech production constrains perception of verbal transformations (Sato et al., 2006) further implicates speech-specific (e.g., articulatory) mechanisms being important for generating verbal transformations, as does neurophysiological activity in left inferior frontal speech areas associated with transformations (Sato et al., 2004; Kondo and Kashino, 2007; Basirat et al., 2008).

In addition to speech perception paradigms, signs of multistable perception have also been observed in a variety of musical tasks (e.g., Deutsch, 1997; Toiviainen and Snyder, 2003; Repp, 2007; Iversen et al., 2009). Additional research on musical multistability would be especially interesting in light of evidence suggesting distinct mechanisms for resolving ambiguous stimuli in vision vs. hearing and speech-specific mechanisms in verbal transformations. For instance, it would be important to determine whether different types of ambiguous auditory stimuli (e.g., speech vs. music) are resolved in distinct neural circuits. This would suggest that multistability is controlled not by centralized mechanisms in only a few brain areas but rather by the normal dynamics that are available throughout the cerebral cortex or other brain areas.

FROM SOUNDS TO CONSCIOUS PERCEPTS, OR NOT

While the research described above demonstrates the promise of using segregation paradigms to understand the role of high-level factors in resolving ambiguous stimuli, another important topic is to understand why some auditory stimuli fail to become accessible to awareness in the first place. Fortunately, researchers have developed a number of clever techniques, often inspired by similar research in vision, to manipulate whether an auditory event is made consciously accessible to observers. Such techniques are critical to understand the mechanisms underlying stimulus awareness,

and also evaluating the influence of being aware of a stimulus on processing subsequent stimuli, separate from the influence of other factors such as attention (Koch and Tsuchiya, 2007).

ENERGY TRADING

Traditional ASA theory (Bregman, 1990) makes a common, but perhaps erroneous, assumption of the existence of energy trading. According to the energy trading hypothesis, if one auditory component contributes to two objects simultaneously, then the total energy in that component should be split between the two objects so that the sum of the amount of energy the component contributes to each object equals the total amount of energy in the component. Research on this topic provides important insights about how low-level sound components contribute to perception of auditory objects and streams. However, the object representations in a scene do not always split the total amount of energy available in a zero-sum fashion (Shinn-Cunningham et al., 2007). In this study, a pure-tone target was used that could be perceptually grouped with either a rhythmic sequence of pure tones of the same frequency (tone sequence) or with concurrent pure tones of different frequencies (a vowel). If the target was incorporated into the vowel, the category of the vowel would change from /I/ to /ɛ/, and if the target was incorporated into the sequence, its rhythm would change from "galloping" to "even." The tone sequence, vowel, and target were presented together with varying spatial configurations. The target could be presented at the same spatial location as the vowel (or tone sequence) to increase the probability of perceptual grouping, or the target could be presented at a different spatial location. The authors conducted trials in which listeners attended to the vowel while ignoring the tone sequence or vice versa.

They found that in the attend-tone block, listeners heard the target as contributing to the tone sequence in all spatial configurations, except when the target was presented at the same location as the vowel. Oddly, in the attend-vowel block, when the feature was presented at the same spatial location as the vowel, the feature did not group with the vowel – the vowel was perceived as /I/. Because the target did not contribute to either percept (the tone sequence or the vowel), it was as if the target tone disappeared from the mixture. This curious case of a feature disappearing suggests that energy trading does not always hold between objects in scenes and that there can be sounds in a scene that do not reach conscious perception even though they are otherwise audible. Shinn-Cunningham et al. (2007) further suggest that listeners require more evidence to allocate an auditory component to a sound in a perceptual figure than to reject it to the auditory ground. It should be noted that in two other studies (Shinn-Cunningham et al., 2008; Shinn-Cunningham and Schwartz, 2010), the same researchers used a simultaneous tone complex rather than a vowel as a competing sound with the tone sequence, and found results that were more consistent with energy trading (see also Leung et al., 2011). However, these two studies also used a richer harmonic target sound, which changed the perceived pitch of the tone complex when the target was integrated.

Another line of research that is problematic for the energy trading hypothesis is the well-established finding of duplex perception: an auditory component can contribute to two sounds at the same time (Rand, 1974; Fowler and Rosenblum, 1990). Duplex perception was first demonstrated by Rand (1974). In this study, the second and third formant transitions from a syllable, e.g., "da," were presented to one ear while the rest of the syllable (i.e., the first formant and the remaining second and third formants) was presented to the other ear. This stimulus generated two simultaneous percepts: listeners reported hearing a fully intact syllable in one ear and a non-speech chirp-like sound in the other ear. The identity ("da" vs. "ga") of the syllable was determined by the third formant transition. Even though the critical feature for identification of the syllable was presented at a separate spatial location from the rest of the syllable, the feature was integrated with the other components to create a coherent, identifiable percept (while at the same time creating the separate percept of a chirp).

Duplex perception has been found to be surprisingly resistant to a variety of other manipulations of the third formant transition, such as SOA (e.g., Bentin and Mann, 1990; Nygaard and Eimas, 1990; Nygaard, 1993), amplitude differences (Cutting, 1976; Whalen and Liberman, 1987; Bentin and Mann, 1990), f_0 (Cutting, 1976), and periodicity differences (Repp and Bentin, 1984). The effect is so strong that it has even been found to occur when the isolated formant transition is not necessary to form a coherent percept (Nygaard and Eimas, 1990). Duplex perception phenomena are not limited to speech objects. For example, when two simultaneous piano notes are presented to one ear while a single note is presented simultaneously to the other ear, the resulting percept is of both the single tone and a fused chord (Pastore et al., 1983). Duplex perception also has been demonstrated with environmental sounds (see Fowler and Rosenblum, 1990).

In summary, it is necessary to either modify ASA theory (Bregman, 1990) or to look beyond it for an explanation of the nonveridical perceptual organization of auditory scenes. Collectively, the findings of duplex perception and the recent case of feature non-allocation contradict the energy trading hypothesis and call into question the amount of low-level detail we are aware of in our acoustic environment (cf. Nahum et al., 2008). Future research on energy trading using denser and more naturalistic auditory scenes is needed to provide a more complete picture of how ASA is accomplished to generate our conscious perception of auditory objects and streams.

CHANGE DEAFNESS

Change deafness is the surprising failure to notice striking changes to auditory scenes. A visual analog to this phenomenon has been extensively studied in the visual domain, where it is referred to as change blindness (for reviews, see Rensink, 2002; Simons and Rensink, 2005). And a related auditory phenomenon 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 (Vitevitch, 2003; Sinnett et al., 2006). Studies using the one-shot technique, in which presentation of a scene is followed by an interruption and then either the same or a modified scene, have been the most common way of examining change deafness. Listeners were found to often miss changes to environmental objects, such as a dog barking changing to a piano tune (e.g., Eramudugolla et al., 2005; Gregg and Samuel, 2008, 2009). It is important to note

that change deafness occurs even though scenes sizes are typically quite small: ∼45% change deafness occurred in Gregg and Samuel (2008) with just four objects per scene. An understanding of the mechanisms underlying change deafness has the potential to inform several issues in auditory perception, such as the completeness of our representation of the auditory world, the limitations of the auditory perceptual system, and how auditory perception may limit auditory memory for objects (for a review, see Snyder and Gregg, 2011). Change deafness might also be useful for studying unconscious processing of changes, as well as the mechanisms that enable changes to reach awareness.

One study has shown that change deafness is reduced with directed attention to the changing object (Eramudugolla et al., 2005). In this study, a 5-s scene was presented, followed by a burst of white noise, and then another 5 s scene that was either the same or different. On Different trials, an object from Scene 1 was either 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," and substantial change deafness was found when not attending to the to-be-changed object. However, when attention was directed to the to-be-changed object via a verbal cue, change detection performance was nearly perfect. One problem with this study, however, is that attention cues were always valid. As a result, participants could have listened for the cued sound in Scene 2, rather than actually comparing the two scenes. An interesting question to address in future research is what aspects of auditory objects must be attended to enhance performance.

Failures to detect changes may not necessarily reflect a failure to encode objects in scenes. Gregg and Samuel (2008) presented an auditory scene, followed by a burst of noise, and then another scene that was either the same as or different than the first scene. Participants performed a change detection task, followed by an object-encoding task, in which they indicated which of two objects they had heard in one of the two scenes. Gregg and Samuel found that object-encoding had a lower error rate than change detection (28 vs. 53%). This study also found that the acoustics of a scene were a critical determinant of change deafness: performance improved when the object that changed was more acoustically distinct from the sound it replaced. 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 suggested that successful change detection may not be based on object identification, as is traditionally 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.

Recently, however, McAnally et al. (2010) distinguished between object-encoding on detected and not detected change trials and found that performance in identifying which object was deleted was near ceiling when changes were detected but at chance when changes were not detected. This finding suggests that changes may only be detected if objects are well encoded, contrary to the findings of Gregg and Samuel (2008). However, it should be noted that the extent of change deafness that occurred in McAnally et al. (2010) was quite modest. They obtained 15% change deafness for scene sizes of four objects, whereas Gregg and Samuel obtained 45% change deafness for scene sizes of four objects. One potential reason for the discrepancy across studies may be that the task in McAnally et al. (2010) did not elicit much change deafness. In their study, a changed scene consisted of an object that was missing, rather than an object replaced by a different object as in Gregg and Samuel. Despite the task differences, the results of McAnally et al. (2010) do question the extent to which objects are encoded during change deafness, and this is an issue that warrants further investigation.

One major issue in the change deafness research is the question of whether change deafness actually reflects verbal or semantic processing limitations, rather than a sensory-level process. Gregg and Samuel (2009) have shown that abstract identity information seems to be encoded preferentially compared to intricate physical detail. 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. 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 compared to between-category changes. Gregg and Samuel did not address the specific nature of the high-level representation being used; it is possible that subjects may have been forming a mental list of verbal labels for all of the objects in the pre-change scene, as has been suggested (Demany et al., 2008). Alternatively, higher-order representations might be activated that reflect the semantic similarity between objects within and between categories.

In summary, change deafness is a relatively new and intriguing line of research. Future research is needed to resolve theoretical issues about why failures to detect auditory changes occur. For example, the issue still remains to what extent sensory-related, attention, memory, or comparison processes are responsible for failures to detect changes and how the interaction of these processes contributes to change deafness.

MASKING

Masking of a target stimulus by another stimulus presented around the same time has been used extensively to study low-level mechanisms of auditory processing. Typically, masking has been observed most strongly when the target and masking stimuli are similar in acoustic features such as frequency, which can be attributed to interference in early frequency-specific stages of processing (e.g., Moore, 1978). This form of masking is referred to as "energetic masking," in contrast to "informational masking," which is assumed to occur when sounds do not have acoustic overlap. Rather, informational masking is assumed to take place at later anatomical sites in the auditory system and to result from a variety of higher-level factors including perceptual grouping and attention (Durlach et al., 2003a; Kidd et al., 2007; Shinn-Cunningham, 2008). The notion of informational masking has generated interesting research that can inform perceptual mechanisms relevant to the current discussion. In particular, a variant of the multi-tone masker paradigm (see **Figure 5**) bears some similarity to streaming paradigms in its use of repeating pure tones (Neff and Green,

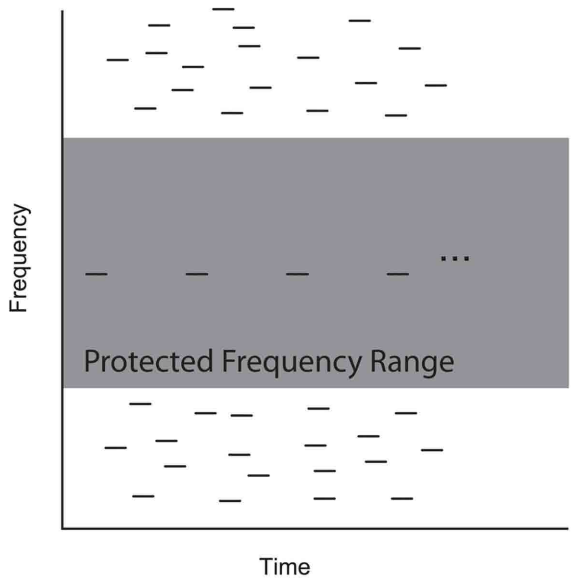

FIGURE 5 | In informational masking experiments, presenting a series of fixed-frequency target tones in the midst of a multi-tone masker stimulus can prevent awareness of the target, even when the masker tones are prevented from overlapping in frequency with the target by using a protected frequency range.

1987). An important difference, however, is the fact that the task typically used in informational masking experiments is to detect whether a fixed-frequency tone is present or absent in a scene along with numerous other masking tones of different frequencies. Peripheral masking can be prevented by not presenting any of the masking tones within a critical band around the target tone.

Several results in the literature have demonstrated interesting similarities between factors that cause streaming and factors that cause release from informational masking. In particular, faster presentation rate, greater target-mask dissimilarity, and cueing the location of the target all facilitate release from masking (Kidd et al., 1994, 2003, 2005; Durlach et al., 2003b; Micheyl et al., 2007b). Similarities may also exist at the neural level: in one study a long-latency response from secondary auditory cortex occurred in response to target tones in a multi-tone masker, but only when participants detected them; remarkably, when the tones were not detected all long-latency brain responses were conspicuously absent (Gutschalk et al., 2008). The response was referred to as an awareness-related negativity (ARN) and was later in latency than (but had similar scalp distribution to) the well-studied N1 response (Näätänen and Picton, 1987), which is consistent with the involvement of negative long-latency responses in streaming and concurrent sound segregation (e.g., Alain et al., 2001; Gutschalk et al., 2005; Snyder et al., 2006). Activity from primary auditory cortex was present regardless of whether the target was detected, strongly suggesting that neural activity must reach beyond primary auditory cortex in order to generate perception. The results were also consistent with the reduction in the N1 observed when sounds are ignored and during sleep (Crowley and Colrain, 2004). The N1 is thought to be an obligatory stimulus-driven response, but if the ARN were related to the N1 (as was suggested by similar source locations), this study would be the first to demonstrate that the N1 generators require participants to be aware of a stimulus to be activated. However, some caution is warranted because the ARN was found to have a longer latency than is typical of the N1, and could therefore be more related to a later negative wave (Nd), which is linked to selective attention (Hansen and Hillyard, 1980). This raises the possibility that the ARN could simply be an index of fluctuations in attention, rather than a direct correlate of awareness.

These results are interesting to compare with findings from a single patient with bilateral superior temporal auditory cortex lesions due to stroke, who performed well on sound detection tasks as long as attention was paid to the tasks (Engelien et al., 2000). However, it is not totally clear what the exact experience of this patient was. In particular, the patient may have had normal conscious experience of detecting sounds as long as enough attention was used; alternatively, the patient may have had little conscious experience of the sounds that he was nevertheless able to reliably detect, in an analogous fashion to patients with blindsight as a result of visual cortex damage (e.g., Stoerig and Cowey, 1997). The same patient showed activation during attention to auditory tasks in a number of brain areas, measured by positron emission tomography, such as in the prefrontal and middle temporal cortices, caudate nucleus, putamen, thalamus, and the cerebellum. Thus, detection of sounds (whether accompanied by conscious experience of the sound or not) may be possible by activating non-auditory brain areas, raising the question of the extent to which superior temporal auditory cortex is necessary or sufficient for awareness to occur. For example, it is possible that the ARN found by Gutschalk et al. (2008) is the result of input from higher-level brain areas that are responsible for generating awareness. Recently, evidence in support of the importance of feedback for generating awareness was found by recording electrophysiological responses in patients in a vegetative state, who compared to controls showed a lack of functional connectivity from frontal to temporal cortex during processing of changes in pure-tone frequency (Boly et al., 2011; for evidence from the visual domain supporting the importance of top-down feedback for perceptual awareness in fully awake, non-brain-damaged individuals, see Pascual-Leone and Walsh, 2001; Wibral et al., 2009).

SUBLIMINAL SPEECH

Recently, researchers have made speech inaudible to determine the extent of auditory priming that can occur without awareness of the priming stimulus. This is one of the only examples of research that has addressed the necessity or sufficiency of auditory awareness for prior stimuli to influence later processing. In one study, priming words were made inaudible by attenuation, time-compression, and masking with time reversals of other time-compressed words immediately before and after the priming words (Kouider and Dupoux, 2005). Compressing words so they were as short as 35 or 40% of their original duration led to very little awareness of the primes as measured on independent tests in which participants had to decide whether the masked sound was a word vs. non-word or a word vs. reversed word. The test word, which was not attenuated or compressed, was played immediately after the

priming word (and simultaneously with the post-priming mask). Non-word pairs were also used that were the same or acoustically similar. Based on the speed with which participants made word vs. non-word decisions about the target, this study showed that repetition of the same word caused priming (i.e., faster responses compared to unrelated prime–target pairs) at all time compressions, including ones that made the prime inaudible (35 and 40%), although the priming effect was larger for audible primes (50 and 70%). Priming also occurred when the prime and target were the same words spoken by different-gender voices, even for the 35% compression level, suggesting that subliminal priming can occur at the abstract word level, independent of the exact acoustics of the sound. Priming effects did not occur for non-words or for semantically related (but acoustically different) words at the subliminal compression levels, suggesting that semantic processing may require conscious perception of words.

A second study used primes that were compressed by 35%, but this time the researchers made the prime audible on some trials by presenting them with a different inter-aural time difference compared to the masking sounds (Dupoux et al., 2008). Again, word priming only occurred for masked words but not masked non-words; priming occurred for both words and non-words when unmasked; and priming was larger for unmasked compared to masked sounds. Additionally, priming did not decline with longer prime–target delays for unmasked words, but the effect declined rapidly for masked sounds over the course of 1000 ms, suggesting a qualitatively different type of robust memory storage for audible sounds.

The basic masked priming effect was recently confirmed by a separate group, who additionally showed that priming occurs mainly for targets with few phonological neighbors (Davis et al., 2010). But a recent study found semantic priming using auditory prime–target word pairs (Daltrozzo et al., 2011), which was in contrast to the study by Kouider and Dupoux (2005). However, the more recent study showing semantic priming used very low-intensity primes that were not possible to categorize, instead of also using time-compression and masking, which could account for the discrepant findings.

Kouider et al. (2010) recently performed an fMRI study using their masking paradigm. They showed priming-related suppression of activity which may prevent processing of stimuli that have already been presented (Schacter et al., 2004). Decrease in activity was found in the left superior temporal auditory cortex (including Heschl's gyrus and planum temporale) for within-gender word pairs and cross-gender word pairs, and in the right insula for within-gender word pairs. For non-words, a different pattern of activity decrease was found in the frontal lobe and caudate nucleus, in addition to response enhancement in the superior temporal cortex. The function of the brain changes should be interpreted cautiously, however, because the magnitudes of activity decrease did not correlate with the magnitudes of behavioral priming. Nevertheless, the results do show that information about unconscious auditory stimuli can reach fairly high levels of processing, with the particular brain areas involved being dependent on the familiarity or meaningfulness of the stimuli.

Speech-priming results are also interesting to compare with a recent study that used fMRI to measure acoustic sentence processing in individuals who were fully awake, lightly sedated, or deeply sedated (Davis et al., 2007). Neural activity in temporal and frontal speech-processing areas continued to differentiate sentences from matched noise stimuli in light sedation and superior temporal responses continued in deep sedation. In contrast, neural activity did not distinguish sentences with vs. without semantically ambiguous words, consistent with the lack of semantic priming observed by Kouider and Dupoux (2005) but inconsistent with the study by Daltrozzo et al. (2011).

AUDITORY ATTENTIONAL BLINK

Attentional blink (AB) refers to a phenomenon where the correct identification of a first target (T1) impairs the processing of a second target (T2) when presented within several hundred millisecond after T1 (e.g., Broadbent and Broadbent, 1987; Raymond et al., 1992; Chun and Potter, 1995). Although the AB has been studied primarily in the visual modality, there is some evidence to suggest that AB also occurs in the auditory modality (e.g., Duncan et al., 1997; Soto-Faraco et al., 2002; Tremblay et al., 2005; Vachon and Tremblay, 2005; Shen and Mondor, 2006).

In a typical auditory AB experiment, participants are presented with a rapid sequence of auditory stimuli. Participants indicate whether they heard T1 and T2 by sequentially pressing two different buttons. The magnitude of the auditory AB decreases with increased time between successive stimuli (Shen and Mondor, 2006; Shen and Alain, 2010). This is not surprising as the increased SOA provides more time to process T1. Increasing the salience of T1 also increased the auditory AB (Horváth and Burgyan, 2011), most likely by momentarily capturing attention, thereby causing a processing deficit for T2. This effect is short lived and is present only when T2 immediately follows T1.

The auditory AB shows a monotonic linear improvement in performance as a function of the interval between the target and the probe (Shen and Mondor, 2006, 2008). This pattern differs from that observed in the visual modality, where the time course of AB has a U-shaped pattern (Raymond et al., 1992). Based on such findings, some researchers have argued that, instead of reflecting the limitation of attention, the auditory AB may simply reflect the cost of task-switching from searching for T1 to searching for T2 (Potter et al., 1998; Chun and Potter, 2001). However, Shen and Mondor (2006) have argued that the difference between a linear and a U-shaped curve occurs because the auditory system has better temporal resolution than the visual system, allowing T1 and T2 to be processed sequentially rather than simultaneously as in the visual modality.

There is increasing evidence that visual AB is sensitive to attentional allocation during the rapid serial visual presentation task. For instance, providing information about the occurrence of the second target attenuates the visual AB (e.g., Martens and Johnson, 2005; Nieuwenstein, 2006; Zhang et al., 2008; Du and Abrams, 2010). These studies indicate that attention can be directed toward the probe feature or temporal position, thereby facilitating its detection. The findings from these studies also suggest that there is some degree of flexibility in the allocation of processing resources despite the existence of processing bottlenecks (Kahneman, 1973).

There is evidence that auditory AB, like visual AB, can also be modulated by attention orienting in a rapid serial auditory

presentation task (Shen and Alain, 2011). Temporal attention was manipulated by task instruction and was blocked for an entire recording session to emphasize focused attention at the designated temporal position. Behavioral results showed a typical pattern of auditory AB. The hit rate at each position when T2 was attended was significantly higher than the false alarm rate at the same position, indicating that participants were able to temporally allocate their attention to the designated temporal position. The latency of the P3b wave elicited by T2 was shortened when attention was oriented to the designated temporal position. More importantly, the electrophysiological results suggest that the initiating of short-term consolidation for T2 was facilitated when attention was oriented to the designated temporal position.

SUMMARY AND CONCLUSION

The research reviewed here clearly shows that much progress has been made in understanding the influence that high-level factors can have on conscious auditory perception and that tasks developed to study ASA are well suited for such investigations. There is now compelling evidence that one or more attention-related factors (i.e., sustained attention to an entire auditory scene, selective attention to particular objects or streams within a scene, attention switching, attention limitations) can have dramatic influences on perceptual organization of scenes and the ability to detect important events in the environment. However, there is evidence that other high-level mental processes, such as intention and previous knowledge, also greatly impact auditory perception. In some cases, these processes are likely to affect perception indirectly through attention-related processes and in other cases to have more direct influences, although a limitation of the studies reviewed here is that they did not compare effects of attention to other high-level influences to see if they indeed have dissociable effects on perception. In vision studies, one process that has been identified as having distinct influences from attention is the awareness of prior stimuli on perception of subsequent stimuli (Koch and Tsuchiya, 2007). Thus, future studies should attempt to manipulate awareness of prior stimuli, in addition to the other high-level factors discussed here, and directly compare their effects. Fortunately, researchers have made substantial progress in manipulating awareness of acoustic stimuli, as we reviewed here, which should make such efforts possible in the future. However, much additional research is needed to fully understand the mechanisms that mediate various types of auditory awareness, such as measured by informational masking, change deafness, priming, and AB paradigms. More work on patients with damage to auditory brain areas, in addition to neurostimulation, neurophysiology, and computational studies are also likely to reveal important findings about the mechanisms of auditory awareness.

ACKNOWLEDGMENTS

The writing of this paper was supported by National Science Foundation grant BCS1026023 and a summer research stipend from the College of Liberal Arts at the University of Nevada Las Vegas awarded to Joel S. Snyder, and a Canadian Institute for Health Research grant awarded to Claude Alain.

REFERENCES

Ahveninen, J., Hamalainen, M., Jaaskelainen, I. P., Ahlfors, S. P., Huang, S., Lin, F. H., Raij, T., Sams, M., Vasios, C. E., and Belliveau, J. W. (2011). Attention-driven auditory cortex short-term plasticity helps segregate relevant sounds from noise. *Proc. Natl. Acad. Sci. U.S.A.* 108, 4182–4187.

Alain, C. (2007). Breaking the wave: effects of attention and learning on concurrent sound perception. *Hear. Res.* 229, 225–236.

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

Alain, C., Arnott, S. R., and Picton, T. W. (2001). Bottom-up and top-down influences on auditory scene analysis: evidence from event-related brain potentials. *J. Exp. Psychol. Hum. Percept. Perform.* 27, 1072–1089.

Alain, C., and Izenberg, A. (2003). Effects of attentional load on auditory scene analysis. *J. Cogn. Neurosci.* 15, 1063–1073.

Alain, C., Reinke, K., He, Y., Wang, C. H., and Lobaugh, N. (2005). Hearing two things at once: neurophysiological indices of speech segregation and identification. *J. Cogn. Neurosci.* 17, 811–818.

Alain, C., and Woods, D. L. (1993). Distractor clustering enhances detection speed and accuracy during selective listening. *Percept. Psychophys.* 54, 509–514.

Alain, C., and Woods, D. L. (1994). Signal clustering modulates auditory cortical activity in humans. *Percept. Psychophys.* 56, 501–516.

Andreou, L. V., Kashino, M., and Chait, M. (2011). The role of temporal regularity in auditory segregation. *Hear. Res.* 280, 228–235.

Assmann, P. F., and Summerfield, Q. (1990). Modeling the perception of concurrent vowels: vowels with different fundamental frequencies. *J. Acoust. Soc. Am.* 88, 680–697.

Basirat, A., Sato, M., Schwartz, J. L., Kahane, P., and Lachaux, J. P. (2008). Parieto-frontal gamma band activity during the perceptual emergence of speech forms. *Neuroimage* 42, 404–413.

Bendixen, A., Denham, S. L., Gyimesi, K., and Winkler, I. (2010). Regular patterns stabilize auditory streams. *J. Acoust. Soc. Am.* 128, 3658–3666.

Bentin, S., and Mann, V. (1990). Masking and stimulus-intensity effects on duplex perception: a confirmation of the dissociation between speech and nonspeech modes. *J. Acoust. Soc. Am.* 88, 64–74.

Bey, C., and McAdams, S. (2002). Schema-based processing in auditory scene analysis. *Percept. Psychophys.* 64, 844–854.

Bidet-Caulet, A., Fischer, C., Besle, J., Aguera, P. E., Giard, M. H., and Bertrand, O. (2007). Effects of selective attention on the electrophysiological representation of concurrent sounds in the human auditory cortex. *J. Neurosci.* 27, 9252–9261.

Biederman, I. (1987). Recognition-by-components: a theory of human image understanding. *Psychol. Rev.* 94, 115–147.

Boly, M., Garrido, M. I., Gosseries, O., Bruno, M. A., Boveroux, P., Schnakers, C., Massimini, M., Litvak, V., Laureys, S., and Friston, K. (2011). Preserved feedforward but impaired top-down processes in the vegetative state. *Science* 332, 858–862.

Boring, E. G. (1953). A history of introspection. *Psychol. Bull.* 50, 169–189.

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

Bregman, A. S., and Campbell, J. (1971). Primary auditory stream segregation and perception of order in rapid sequences of tones. *J. Exp. Psychol.* 89, 244–249.

Broadbent, D. E., and Broadbent, M. H. (1987). From detection to identification: response to multiple targets in rapid serial visual presentation. *Percept. Psychophys.* 42, 105–113.

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

Carlyon, R. P., Cusack, R., Foxton, J. M., and Robertson, I. H. (2001). Effects of attention and unilateral neglect on auditory stream segregation. *J. Exp. Psychol. Hum. Percept. Perform.* 27, 115–127.

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

Chalikia, M. H., and Bregman, A. S. (1989). The perceptual segregation of simultaneous auditory signals: pulse train segregation and vowel segregation. *Percept. Psychophys.* 46, 487–496.

Cherry, E. C. (1953). Some experiments on the recognition of speech, with one and with two ears. *J. Acoust. Soc. Am.* 25, 975–979.

Chun, M. M., and Potter, M. C. (1995). A two-stage model for multiple target detection in rapid serial visual presentation. *J. Exp. Psychol. Hum. Percept. Perform.* 21, 109–127.

Chun, M. M., and Potter, M. C. (2001). "The attentional blink and task switching within and across modalities," in *The Limits of Attention: Temporal Constraints in Human Information Processing*, ed. K. Shapiro (Oxford: Oxford University Press), 20–35.

Ciocca, V. (2008). The auditory organization of complex sounds. *Front. Biosci.* 13, 148–169.

Crick, F., and Koch, C. (1995). Are we aware of neural activity in primary visual cortex? *Nature* 375, 121–123.

Crick, F., and Koch, C. (2003). A framework for consciousness. *Nat. Neurosci.* 6, 119–126.

Crowley, K. E., and Colrain, I. M. (2004). A review of the evidence for P2 being an independent component process: age, sleep and modality. *Clin. Neurophysiol.* 115, 732–744.

Cusack, R. (2005). The intraparietal sulcus and perceptual organization. *J. Cogn. Neurosci.* 17, 641–651.

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

Cutting, J. E. (1976). Auditory and linguistic processes in speech-perception: inferences from 6 fusions in dichotic-listening. *Psychol. Rev.* 83, 114–140.

Daltrozzo, J., Signoret, C., Tillmann, B., and Perrin, F. (2011). Subliminal semantic priming in speech. *PLoS ONE* 6, e20273. doi:10.1371/journal.pone.0020273

Danzinger, K. (1980). The history of introspection reconsidered. *J. Hist. Behav. Sci.* 16, 241–262.

Davis, C., Kim, J., and Barbaro, A. (2010). Masked speech priming: neighborhood size matters. *J. Acoust. Soc. Am.* 127, 2110–2113.

Davis, M. H., Coleman, M. R., Absalom, A. R., Rodd, J. M., Johnsrude, I. S., Matta, B. F., Owen, A. M., and Menon, D. K. (2007). Dissociating speech perception and comprehension at reduced levels of awareness. *Proc. Natl. Acad. Sci. U.S.A.* 104, 16032–16037.

Demany, L., Trost, W., Serman, M., and Semal, C. (2008). Auditory change detection: simple sounds are not memorized better than complex sounds. *Psychol. Sci.* 19, 85–91.

Denham, S. L., and Winkler, I. (2006). The role of predictive models in the formation of auditory streams. *J. Physiol. Paris* 100, 154–170.

Deutsch, D. (1997). The tritone paradox: a link between music and speech. *Curr. Dir. Psychol. Sci.* 6, 174–180.

Devergie, A., Grimault, N., Tillmann, B., and Berthommier, F. (2010). Effect of rhythmic attention on the segregation of interleaved melodies. *J. Acoust. Soc. Am.* 128, EL1–EL7.

Ding, N., and Simon, J. Z. (2012). Neural coding of continuous speech in auditory cortex during monaural and dichotic listening. *J. Neurophysiol.* 107, 78–89.

Ditzinger, T., Tuller, B., Haken, H., and Kelso, J. A. S. (1997a). A synergetic model for the verbal transformation effect. *Biol. Cybern.* 77, 31–40.

Ditzinger, T., Tuller, B., and Kelso, J. A. S. (1997b). Temporal patterning in an auditory illusion: the verbal transformation effect. *Biol. Cybern.* 77, 23–30.

Donner, T. H., Sagi, D., Bonneh, Y. S., and Heeger, D. J. (2008). Opposite neural signatures of motion-induced blindness in human dorsal and ventral visual cortex. *J. Neurosci.* 28, 10298–10310.

Dowling, W. J. (1973). Perception of interleaved melodies. *Cogn. Psychol.* 5, 322–337.

Dowling, W. J., Lung, K. M. T., and Herrbold, S. (1987). Aiming attention in pitch and time in the perception of interleaved melodies. *Percept. Psychophys.* 41, 642–656.

Du, F., and Abrams, R. A. (2010). Endogenous orienting is reduced during the attentional blink. *Exp. Brain Res.* 205, 115–121.

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

Dupoux, E., De Gardelle, V., and Kouider, S. (2008). Subliminal speech perception and auditory streaming. *Cognition* 109, 267–273.

Durlach, N. I., Mason, C. R., Kidd, G. Jr., Arbogast, T. L., Colburn, H. S., and Shinn-Cunningham, B. G. (2003a). Note on informational masking. *J. Acoust. Soc. Am.* 113, 2984–2987.

Durlach, N. I., Mason, C. R., Shinn-Cunningham, B. G., Arbogast, T. L., Colburn, H. S., and Kidd, G. Jr. (2003b). Informational masking: counteracting the effects of stimulus uncertainty by decreasing target-masker similarity. *J. Acoust. Soc. Am.* 114, 368–379.

Dykstra, A. R., Halgren, E., Thesen, T., Carlson, C. E., Doyle, W., Madsen, J. R., Eskandar, E. N., and Cash, S. S. (2011). Widespread brain areas engaged during a classical auditory streaming task revealed by intracranial EEG. *Front. Hum. Neurosci.* 5:74. doi:10.3389/fnhum.2011.00074

Dyson, B. J., Alain, C., and He, Y. (2005). Effects of visual attentional load on low-level auditory scene analysis. *Cogn. Affect. Behav. Neurosci.* 5, 319–338.

Edelman, S. (1998). Representation is representation of similarities. *Behav. Brain Sci.* 21, 449–498.

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

Elhilali, M., Xiang, J. J., Shamma, S. A., and Simon, J. Z. (2009b). Interaction between attention and bottom-up saliency mediates the representation of foreground and background in an auditory scene. *PLoS Biol.* 7, e1000129. doi:10.1371/journal.pbio.1000129

Engelien, A., Huber, W., Silbersweig, D., Stern, E., Frith, C. D., Doring, W., Thron, A., and Frackowiak, R. S. (2000). The neural correlates of 'deaf-hearing' in man: conscious sensory awareness enabled by attentional modulation. *Brain* 123, 532–545.

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

Fowler, C. A., and Rosenblum, L. D. (1990). Duplex perception: a comparison of monosyllables and slamming doors. *J. Exp. Psychol. Hum. Percept. Perform.* 16, 742–754.

Greene, M. R., and Oliva, A. (2009). Recognition of natural scenes from global properties: seeing the forest without representing the trees. *Cogn. Psychol.* 58, 137–176.

Gregg, M. K., and Samuel, A. G. (2008). Change deafness and the organizational properties of sounds. *J. Exp. Psychol. Hum. Percept. Perform.* 34, 974–991.

Gregg, M. K., and Samuel, A. G. (2009). The importance of semantics in auditory representations. *Atten. Percept. Psychophys.* 71, 607–619.

Gutschalk, A., Micheyl, C., Melcher, J. R., Rupp, A., Scherg, M., and Oxenham, A. J. (2005). Neuromagnetic correlates of streaming in human auditory cortex. *J. Neurosci.* 25, 5382–5388.

Gutschalk, A., Micheyl, C., and Oxenham, A. J. (2008). Neural correlates of auditory perceptual awareness under informational masking. *PLoS Biol.* 6, e138. doi:10.1371/journal.pbio.0060138

Hansen, J. C., and Hillyard, S. A. (1980). Endogenous brain potentials associated with selective auditory attention. *Electroencephalogr. Clin. Neurophysiol.* 49, 277–290.

Hautus, M. J., and Johnson, B. W. (2005). Object-related brain potentials associated with the perceptual segregation of a dichotically embedded pitch. *J. Acoust. Soc. Am.* 117, 275–280.

Hill, K. T., Bishop, C. W., Yadav, D., and Miller, L. M. (2011). Pattern of BOLD signal in auditory cortex relates acoustic response to perceptual streaming. *BMC Neurosci.* 12, 85. doi:10.1186/1471-2202-12-85

Hillyard, S. A., Hink, R. F., Schwent, V. L., and Picton, T. W. (1973). Electrical signs of selective attention in the human brain. *Science* 182, 171–180.

Hochstein, S., and Ahissar, M. (2002). View from the top: hierarchies and reverse hierarchies in the visual system. *Neuron* 36, 791–804.

Hock, H. S., Kelso, J. A. S., and Schoner, G. (1993). Bistability and hysteresis in the organization of apparent motion patterns. *J. Exp. Psychol. Hum. Percept. Perform.* 19, 63–80.

Horváth, J., and Burgyan, A. (2011). Distraction and the auditory attentional blink. *Atten. Percept. Psychophys.* 73, 695–701.

Hupé, J. M., Joffo, L. M., and Pressnitzer, D. (2008). Bistability for audiovisual stimuli: perceptual decision is modality specific. *J. Vis.* 8, 1–15.

Hupé, J. M., and Rubin, N. (2003). The dynamics of bi-stable alternation in ambiguous motion displays: a fresh look at plaids. *Vision Res.* 43, 531–548.

Iversen, J. R., Repp, B. H., and Patel, A. D. (2009). "Top-down control of rhythm perception modulates early auditory responses," in *Neurosciences and Music III: Disorders and Plasticity*, eds S. Dallabella, N. Kraus, K. Overy, C. Pantev, J. S. Snyder, M. Tervaniemi, B. Tillmann, and G. Schlaug (Oxford: Blackwell Publishing), 58–S73.

Johnson, B. W., Hautus, M., and Clapp, W. C. (2003). Neural activity associated with binaural processes for the perceptual segregation of pitch. *Clin. Neurophysiol.* 114, 2245–2250.

Jones, M. R. (1976). Time, our lost dimension – toward a new theory of perception, attention, and memory. *Psychol. Rev.* 83, 323–355.

Jones, M. R., and Boltz, M. (1989). Dynamic attending and responses to time. *Psychol. Rev.* 96, 459–491.

Jones, M. R., Kidd, G., and Wetzel, R. (1981). Evidence for rhythmic attention. *J. Exp. Psychol. Hum. Percept. Perform.* 7, 1059–1073.

Kahneman, D. (1973). *Attention and Effort*. Englewood Cliffs, NJ: Prentice-Hall.

Kanai, R., Tsuchiya, N., and Verstraten, F. A. (2006). The scope and limits of top-down attention in unconscious visual processing. *Curr. Biol.* 16, 2332–2336.

Kerlin, J. R., Shahin, A. J., and Miller, L. M. (2010). Attentional gain control of ongoing cortical speech representations in a "cocktail party." *J. Neurosci.* 30, 620–628.

Kidd, G. Jr., Arbogast, T. L., Mason, C. R., and Gallun, F. J. (2005). The advantage of knowing where to listen. *J. Acoust. Soc. Am.* 118, 3804–3815.

Kidd, G. Jr., Mason, C. R., and Richards, V. M. (2003). Multiple bursts, multiple looks, and stream coherence in the release from informational masking. *J. Acoust. Soc. Am.* 114, 2835–2845.

Kidd, G., Mason, C. R., Deliwala, P. S., Woods, W. S., and Colburn, H. S. (1994). Reducing informational masking by sound segregation. *J. Acoust. Soc. Am.* 95, 3475–3480.

Kidd, G., Mason, C. R., Richards, V. M., Gallun, F. J., and Durlach, N. I. (2007). "Informational masking," in *Auditory Perception of Sound Sources*, eds W. A. Yost, R. R. Fay, and A. N. Popper (New York: Springer), 143–189.

Kim, C. Y., and Blake, R. (2005). Psychophysical magic: rendering the visible 'invisible.' *Trends Cogn. Sci. (Regul. Ed.)* 9, 381–388.

Koch, C., and Tsuchiya, N. (2007). Attention and consciousness: two distinct brain processes. *Trends Cogn. Sci. (Regul. Ed.)* 11, 16–22.

Kondo, H. M., and Kashino, M. (2007). Neural mechanisms of auditory awareness underlying verbal transformations. *Neuroimage* 36, 123–130.

Kondo, H. M., and Kashino, M. (2009). Involvement of the thalamocortical loop in the spontaneous switching of percepts in auditory streaming. *J. Neurosci.* 29, 12695–12701.

Kouider, S., de Gardelle, V., Dehaene, S., Dupoux, E., and Pallier, C. (2010). Cerebral bases of subliminal speech priming. *Neuroimage* 49, 922–929.

Kouider, S., and Dupoux, E. (2005). Subliminal speech priming. *Psychol. Sci.* 16, 617–625.

Lamme, V. A. (2004). Separate neural definitions of visual consciousness and visual attention; a case for phenomenal awareness. *Neural. Netw.* 17, 861–872.

Large, E. W., and Jones, M. R. (1999). The dynamics of attending: how people track time-varying events. *Psychol. Rev.* 106, 119–159.

Leopold, D. A., and Logothetis, N. K. (1999). Multistable phenomena: changing views in perception. *Trends Cogn. Sci. (Regul. Ed.)* 3, 254–264.

Leung, A. W., Jolicoeur, P., Vachon, F., and Alain, C. (2011). The perception of concurrent sound objects in harmonic complexes impairs gap detection. *J. Exp. Psychol. Hum. Percept. Perform.* 37, 727–736.

Long, G. M., and Toppino, T. C. (2004). Enduring interest in perceptual ambiguity: alternating views of reversible figures. *Psychol. Bull.* 130, 748–768.

Martens, S., and Johnson, A. (2005). Timing attention: cuing target onset interval attenuates the attentional blink. *Mem. Cognit.* 33, 234–240.

McAnally, K. I., Martin, R. L., Eramudugolla, R., Stuart, G. W., Irvine, D. R. F., and Mattingley, J. B. (2010). A dual-process account of auditory change detection. *J. Exp. Psychol. Hum. Percept. Perform.* 36, 994–1004.

McClelland, J. L., Mirman, D., and Holt, L. L. (2006). Are there interactive processes in speech perception? *Trends Cogn. Sci. (Regul. Ed.)* 10, 363–369.

McDermott, J. H., Wrobleski, D., and Oxenham, A. J. (2011). Recovering sound sources from embedded repetition. *Proc. Natl. Acad. Sci. U.S.A.* 108, 1188–1193.

McDonald, K. L., and Alain, C. (2005). Contribution of harmonicity and location to auditory object formation in free field: evidence from event-related brain potentials. *J. Acoust. Soc. Am.* 118, 1593–1604.

Micheyl, C., Carlyon, R. P., Gutschalk, A., Melcher, J. R., Oxenham, A. J., Rauschecker, J. P., Tian, B., and Courtenay Wilson, E. (2007a). The role of auditory cortex in the formation of auditory streams. *Hear. Res.* 229, 116–131.

Micheyl, C., Shamma, S. A., and Oxenham, A. J. (2007b). "Hearing out repeating elements in randomly varying multitone sequences: a case of streaming?" in *Hearing: From Sensory Processing to Perception*, eds B. Kollmeier, G. Klump, V. Hohmann, U. Langemann, M. Mauermann, S. Uppenkamp, and J. Verhey. (Berlin: Springer), 267–274.

Micheyl, C., and Oxenham, A. J. (2010). Objective and subjective psychophysical measures of auditory stream integration and segregation. *J. Assoc. Res. Otolaryngol.* 11, 709–724.

Micheyl, C., Tian, B., Carlyon, R. P., and Rauschecker, J. P. (2005). Perceptual organization of tone sequences in the auditory cortex of awake macaques. *Neuron* 48, 139–148.

Miller, G. A., and Heise, G. A. (1950). The trill threshold. *J. Acoust. Soc. Am.* 22, 637–638.

Moore, B. C. J. (1978). Psychophysical tuning curves measured in simultaneous and forward masking. *J. Acoust. Soc. Am.* 63, 524–532.

Moore, B. C. J., Glasberg, B. R., and Peters, R. W. (1986). Thresholds for hearing mistuned partials as separate tones in harmonic complexes. *J. Acoust. Soc. Am.* 80, 479–483.

Moore, B. C. J., and Gockel, H. (2002). Factors influencing sequential stream segregation. *Acta Acust. United Ac.* 88, 320–333.

Movshon, J. A., Adelson, E. H., Gizzi, M. S., and Newsome, W. T. (1985). "The analysis of moving visual patterns," in *Study Group on Pattern Recognition Mechanisms*, eds C. Chagas, R. Gattass, and C. G. Gross (Vatican City: Pontifica Academia Scientiarum), 117–151.

Näätänen, R., and Picton, T. (1987). The N1 wave of the human electric and magnetic response to sound: a review and an analysis of the component structure. *Psychophysiology* 24, 375–425.

Nahum, M., Nelken, I., and Ahissar, M. (2008). Low-level information and high-level perception: the case of speech in noise. *PLoS Biol.* 6, e126. doi:10.1371/journal.pbio.0060126

Neff, D. L., and Green, D. M. (1987). Masking produced by spectral uncertainty with multicomponent maskers. *Percept. Psychophys.* 41, 409–415.

Newman, R. S., and Evers, S. (2007). The effect of talker familiarity on stream segregation. *J. Phon.* 35, 85–103.

Nieuwenstein, M. R. (2006). Top-down controlled, delayed selection in the attentional blink. *J. Exp. Psychol. Hum. Percept. Perform.* 32, 973–985.

Nygaard, L. C. (1993). Phonetic coherence in duplex perception: effects of acoustic differences and lexical status. *J. Exp. Psychol. Hum. Percept. Perform.* 19, 268–286.

Nygaard, L. C., and Eimas, P. D. (1990). A new version of duplex perception: evidence for phonetic and nonphonetic fusion. *J. Acoust. Soc. Am.* 88, 75–86.

Oliva, A., and Torralba, A. (2001). Modeling the shape of the scene: a holistic representation of the spatial envelope. *Int. J. Comput. Vis.* 42, 145–175.

Pascual-Leone, A., and Walsh, V. (2001). Fast backprojections from the motion to the primary visual area necessary for visual awareness. *Science* 292, 510–512.

Pastore, R. E., Schmuckler, M. A., Rosenblum, L., and Szczesiul, R. (1983). Duplex perception with musical stimuli. *Percept. Psychophys.* 33, 469–474.

Pearson, J., and Brascamp, J. (2008). Sensory memory for ambiguous vision. *Trends Cogn. Sci. (Regul. Ed.)* 12, 334–341.

Pichora-Fuller, M. K., Schneider, B. A., and Daneman, M. (1995). How young and old adults listen to and remember speech in noise. *J. Acoust. Soc. Am.* 97, 593–608.

Pitt, M. A., and Shoaf, L. (2002). Linking verbal transformations to their causes. *J. Exp. Psychol. Hum. Percept. Perform.* 28, 150–162.

Potter, M. C., Chun, M. M., Banks, B. S., and Muckenhoupt, M. (1998). Two attentional deficits in serial target search: the visual attentional blink and an amodal task-switch deficit. *J. Exp. Psychol. Learn Mem. Cogn.* 24, 979–992.

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

Pressnitzer, D., Sayles, M., Micheyl, C., and Winter, I. M. (2008). Perceptual organization of sound begins in the auditory periphery. *Curr. Biol.* 18, 1124–1128.

Rand, T. C. (1974). Dichotic release from masking for speech. *J. Acoust. Soc. Am.* 55, 678–680.

Raymond, J. E., Shapiro, K. L., and Arnell, K. M. (1992). Temporary suppression of visual processing in an RSVP task: an attentional blink? *J. Exp. Psychol. Hum. Percept. Perform.* 18, 849–860.

Rensink, R. A. (2002). Change detection. *Annu. Rev. Psychol.* 53, 245–277.

Repp, B. H. (2007). Hearing a melody in different ways: multistability of metrical interpretation, reflected in rate limits of sensorimotor synchronization. *Cognition* 102, 434–454.

Repp, B. H., and Bentin, S. (1984). Parameters of spectral temporal fusion in speech-perception. *Percept. Psychophys.* 36, 523–530.

Riecke, L., Mendelsohn, D., Schreiner, C., and Formisano, E. (2009). The continuity illusion adapts to the

auditory scene. *Hear. Res.* 247, 71–77.

Riecke, L., Micheyl, C., Vanbussel, M., Schreiner, C. S., Mendelsohn, D., and Formisano, E. (2011). Recalibration of the auditory continuity illusion: sensory and decisional effects. *Hear. Res.* 152–162.

Roberts, B., Glasberg, B. R., and Moore, B. C. J. (2002). Primitive stream segregation of tone sequences without differences in fundamental frequency or passband. *J. Acoust. Soc. Am.* 112, 2074–2085.

Rogers, W. L., and Bregman, A. S. (1993). An experimental evaluation of three theories of auditory stream segregation. *Percept. Psychophys.* 53, 179–189.

Sato, M., Baciu, M., Loevenbruck, H., Schwartz, J. L., Cathiard, M. A., Segebarth, C., and Abry, C. (2004). Multistable representation of speech forms: a functional MRI study of verbal transformations. *Neuroimage* 23, 1143–1151.

Sato, M., Schwartz, J. L., Abry, C., Cathiard, M. A., and Loevenbruck, H. (2006). Multistable syllables as enacted percepts: a source of an asymmetric bias in the verbal transformation effect. *Percept. Psychophys.* 68, 458–474.

Schacter, D. L., Dobbins, I. G., and Schnyer, D. M. (2004). Specificity of priming: a cognitive neuroscience perspective. *Nat. Rev. Neurosci.* 5, 853–862.

Schadwinkel, S., and Gutschalk, A. (2011). Transient BOLD activity locked to perceptual reversals of auditory streaming in human auditory cortex and inferior colliculus. *J. Neurophysiol.* 105, 1977–1983.

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

Shen, D., and Alain, C. (2010). Neuroelectric correlates of auditory attentional blink. *Psychophysiology* 47, 184–191.

Shen, D., and Alain, C. (2011). Temporal attention facilitates short-term consolidation during a rapid serial auditory presentation task. *Exp. Brain Res.* 215, 285–292.

Shen, D., and Mondor, T. A. (2006). Effect of distractor sounds on the auditory attentional blink. *Percept. Psychophys.* 68, 228–243.

Shen, D., and Mondor, T. A. (2008). Object file continuity and the auditory attentional blink. *Percept. Psychophys.* 70, 896–915.

Shinn-Cunningham, B. G. (2008). Object-based auditory and visual attention. *Trends Cogn. Sci. (Regul. Ed.)* 12, 182–186.

Shinn-Cunningham, B. G., Lee, A. K. C., and Babcock, S. (2008). Measuring the perceived content of auditory objects using a matching paradigm. *J. Assoc. Res. Otolaryngol.* 9, 388–397.

Shinn-Cunningham, B. G., Lee, A. K. C., and Oxenham, A. J. (2007). A sound element gets lost in perceptual competition. *Proc. Natl. Acad. Sci. U.S.A.* 104, 12223–12227.

Shinn-Cunningham, B. G., and Schwartz, A. H. (2010). Dissociation of perceptual judgments of "what" and "where" in an ambiguous auditory scene. *J. Acoust. Soc. Am.* 128, 3041–3051.

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

Sinnett, S., Costa, A., and Soto-Faraco, S. (2006). Manipulating inattentional blindness within and across sensory modalities. *Q. J. Exp. Psychol.* 59, 1425–1442.

Snyder, J. S., and Alain, C. (2007). Toward a neurophysiological theory of auditory stream segregation. *Psychol. Bull.* 133, 780–799.

Snyder, J. S., Alain, C., and Picton, T. W. (2006). Effects of attention on neuroelectric correlates of auditory stream segregation. *J. Cogn. Neurosci.* 18, 1–13.

Snyder, J. S., Carter, O. L., Hannon, E. E., and Alain, C. (2009a). Adaptation reveals multiple levels of representation in auditory stream segregation. *J. Exp. Psychol. Hum. Percept. Perform.* 35, 1232–1244.

Snyder, J. S., Holder, W. T., Weintraub, D. M., Carter, O. L., and Alain, C. (2009b). Effects of prior stimulus and prior perception on neural correlates of auditory stream segregation. *Psychophysiology* 46, 1208–1215.

Snyder, J. S., Carter, O. L., Lee, S.-K., Hannon, E. E., and Alain, C. (2008). Effects of context on auditory stream segregation. *J. Exp. Psychol. Hum. Percept. Perform.* 34, 1007–1016.

Snyder, J. S., and Gregg, M. K. (2011). Memory for sound, with an ear toward hearing in complex scenes. *Atten. Percept. Psychophys.* 73, 1993–2007.

Snyder, J. S., and Weintraub, D. M. (2011). Pattern specificity in the effect of prior Δf on auditory stream segregation. *J. Exp. Psychol. Hum. Percept. Perform.* 37, 1649–1656.

Soto-Faraco, S., Spence, C., Fairbank, K., Kingstone, A., Hillstrom, A. P., and Shapiro, K. (2002). A crossmodal attentional blink between vision and touch. *Psychon. Bull. Rev.* 9, 731–738.

Stainsby, T. H., Moore, B. C. J., Medland, P. J., and Glasberg, B. R. (2004). Sequential streaming and effective level differences due to phase-spectrum manipulations. *J. Acoust. Soc. Am.* 115, 1665–1673.

Stoerig, P., and Cowey, A. (1997). Blindsight in man and monkey. *Brain* 120, 535–559.

Sussman, E., Ritter, W., and Vaughan, H. G. (1999). An investigation of the auditory streaming effect using event-related brain potentials. *Psychophysiology* 36, 22–34.

Sussman, E. S., Horváth, J., Winkler, I., and Orr, M. (2007). The role of attention in the formation of auditory streams. *Percept. Psychophys.* 69, 136–152.

Teki, S., Chait, M., Kumar, S., Von Kriegstein, K., and Griffiths, T. D. (2011). Brain bases for auditory stimulus-driven figure-ground segregation. *J. Neurosci.* 31, 164–171.

Thompson, S. K., Carlyon, R. P., and Cusack, R. (2011). An objective measurement of the build-up of auditory streaming and of its modulation by attention. *J. Exp. Psychol. Hum. Percept. Perform.* 37, 1253–1262.

Toiviainen, P., and Snyder, J. S. (2003). Tapping to Bach: resonance-based modeling of pulse. *Music Percept.* 21, 43–80.

Tong, F., Meng, M., and Blake, R. (2006). Neural bases of binocular rivalry. *Trends Cogn. Sci. (Regul. Ed.)* 10, 502–511.

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

Tuller, B., Ding, M. Z., and Kelso, J. A. S. (1997). Fractal timing of verbal transforms. *Perception* 26, 913–928.

Ullman, S. (2007). Object recognition and segmentation by a fragment-based hierarchy. *Trends Cogn. Sci. (Regul. Ed.)* 11, 58–64.

Vachon, F., and Tremblay, S. (2005). Auditory attentional blink: masking the second target is necessary, delayed masking is sufficient. *Can. J. Exp. Psychol.* 59, 279–286.

van Ee, R., Van Boxtel, J. J. A., Parker, A. L., and Alais, D. (2009). Multisensory congruency as a mechanism for attentional control over perceptual selection. *J. Neurosci.* 29, 11641–11649.

Van Noorden, L. P. A. S. (1975). *Temporal Coherence in the Perception of Tone Sequences.* Unpublished doctoral dissertation, Eindhoven University of Technology, Eindhoven.

Vitevitch, M. S. (2003). Change deafness: the inability to detect changes between two voices. *J. Exp. Psychol. Hum. Percept. Perform.* 29, 333–342.

Warren, R. M. (1968). Verbal transformation effect and auditory perceptual mechanisms. *Psychol. Bull.* 70, 261–270.

Whalen, D. H., and Liberman, A. M. (1987). Speech-perception takes precedence over nonspeech perception. *Science* 237, 169–171.

Wibral, M., Bledowski, C., Kohler, A., Singer, W., and Muckli, L. (2009). The timing of feedback to early visual cortex in the perception of long-range apparent motion. *Cereb. Cortex* 19, 1567–1582.

Xiang, J. J., Simon, J., and Elhilali, M. (2010). Competing streams at the cocktail party: exploring the mechanisms of attention and temporal integration. *J. Neurosci.* 30, 12084–12093.

Zhang, D., Shao, L., Nieuwenstein, M., and Zhou, X. (2008). Top-down control is not lost in the attentional blink: evidence from intact endogenous cueing. *Exp. Brain Res.* 185, 287–295.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 14 October 2011; paper pending published: 07 November 2011; accepted: 11 January 2012; published online: 07 February 2012.

*Citation: Snyder JS, Gregg MK, Weintraub DM and Alain C (2012) Attention, awareness, and the perception of auditory scenes. Front. Psychology **3**:15. doi: 10.3389/fpsyg.2012.00015*

This article was submitted to Frontiers in Consciousness Research, a specialty of Frontiers in Psychology.