

Toward a Neurophysiological Theory of Auditory Stream Segregation

Joel S. Snyder

Veterans Affairs Boston Healthcare System
and Harvard Medical School

Claude Alain

The Rotman Research Institute

Auditory stream segregation (or streaming) is a phenomenon in which 2 or more repeating sounds differing in at least 1 acoustic attribute are perceived as 2 or more separate sound sources (i.e., streams). This article selectively reviews psychophysical and computational studies of streaming and comprehensively reviews more recent neurophysiological studies that have provided important insights into the mechanisms of streaming. On the basis of these studies, segregation of sounds is likely to occur beginning in the auditory periphery and continuing at least to primary auditory cortex for simple cues such as pure-tone frequency but at stages as high as secondary auditory cortex for more complex cues such as periodicity pitch. Attention-dependent and perception-dependent processes are likely to take place in primary or secondary auditory cortex and may also involve higher level areas outside of auditory cortex. Topographic maps of acoustic attributes, stimulus-specific suppression, and competition between representations are among the neurophysiological mechanisms that likely contribute to streaming. A framework for future research is proposed.

Keywords: auditory scene analysis, attention, auditory cortex, tonotopic organization, neural suppression

Our auditory environment often comprises patterns from multiple sound-producing objects. For example, while talking to a friend in an outdoor urban setting, one might have to separate the friend's speech from the sounds produced by cars passing, birds chirping, an ambulance siren blaring, and a street musician playing an instrument and singing. This is often referred to as the *cocktail party problem* (Cherry, 1953) because a similar situation arises when trying to perceive speech at a noisy social gathering. A type of cocktail party problem also exists for nonhuman animals, for example when frogs, birds, or insects must segregate mating calls from the background noise, and when bats segregate their own echolocation calls from the echoes that return from the environment (Feng & Ratnam, 2000; Hulse, 2002). Because the auditory system does not have an explicit coding of space that unambiguously distinguishes different sound-producing objects, auditory researchers have given much thought to how sounds are segmented and perceptually grouped. The process with which the nervous system makes sense of complex patterns of acoustic stimulation is called *auditory scene analysis* (Bregman, 1990). The goal of auditory scene analysis is to segregate sounds that arise from different environmental sound sources and form internal represen-

tations of auditory objects or streams, such as a friend's speech or another animal's mating call.

Much of the research on auditory scene analysis has occurred following the publication of an article by Al Bregman that introduced the *auditory stream segregation* paradigm (Bregman & Campbell, 1971). This article demonstrated that when pure tones were alternated between a low-frequency range and a high-frequency range with a sufficiently rapid rate, instead of hearing a pattern of alternating low (L) and high (H) tones (i.e., one perceptual stream), participants heard two separate streams of tones, one consisting of L tones and another consisting of H tones (also see Miller & Heise, 1950; Warren, Obusek, Farmer, & Warren, 1969). Bregman and Campbell (1971) demonstrated that when perceiving two segregated streams (known as *streaming* or *fission*), temporal order judgments of tones belonging to different streams became much more difficult than they were when hearing one stream of alternating L and H tones (known as *coherence* or *fusion*). This suggested that the L and H tones were represented separately, preventing comparisons across the two streams of tones. The perception of two streams does not occur instantly following the presentation of two successive events from different frequency ranges. Instead, on a given trial, listeners initially perceive one stream and only after several seconds of buildup does the pattern of alternating tones split into two distinct streams (Anstis & Saida, 1985; Bregman, 1978).

Around the same time as Bregman and Campbell's (1971) article on streaming, Van Noorden (1975) systematically explored the role of stimulus-driven and attention-driven influences on coherence and streaming in his highly influential dissertation. He also introduced a useful stimulus configuration for studying auditory stream segregation that presents L tones (A), H tones (B), and silences (-) in a repeating ABA- pattern (see Figure 1). When the difference in frequency between the A and B tones is small and the repetition rate of the sequence is slow, listeners hear a single stream of tones in a galloping rhythm (i.e., ABA-ABA-). When the

Joel S. Snyder, Department of Psychiatry, Veterans Affairs Boston Healthcare System and Harvard Medical School; Claude Alain, The Rotman Research Institute, Baycrest Centre for Geriatric Care, Toronto, Ontario, Canada.

Joel S. Snyder was supported by National Institutes of Health Clinical Research Training Program Grant T32 MH16259-28 'Clinical Research in Biological and Social Psychiatry.' Claude Alain was supported by the Canadian Institutes of Health Research and the Hearing Foundation of Canada.

Correspondence concerning this article should be addressed to Joel S. Snyder, who is now at the Department of Psychology, University of Nevada, Las Vegas, 4505 S. Maryland Parkway, Box 455030, Las Vegas, NV 89154. E-mail: joel_snyder@hms.harvard.edu

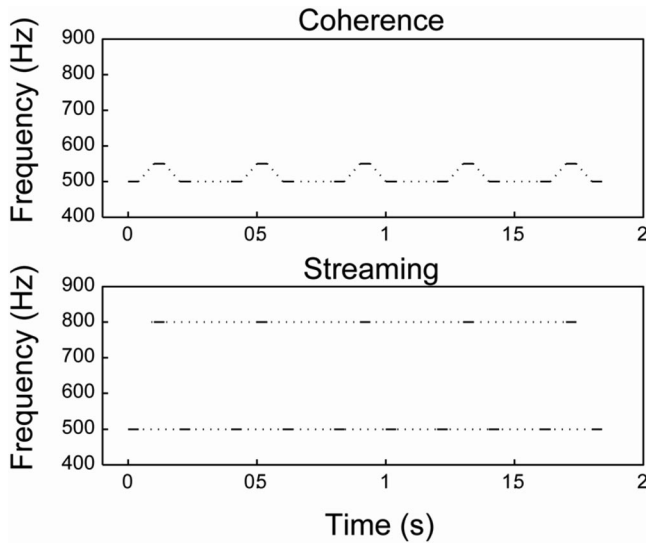


Figure 1. Typical stimulus patterns used in auditory stream segregation experiments, composed of a repeating pattern of low tones (A), high tones (B), and silences (-) in an ABA- pattern. When the A and B tones have similar frequencies, a single stream of tones is heard in a galloping rhythm (known as *coherence*; top panel). When the A and B tones have sufficiently different frequencies, two streams of tones are heard with metronome rhythms (known as *streaming*; bottom panel). In experiments using subjective measures of stream segregation, participants indicate whether they perceive coherence or streaming continuously during the stimulus presentation or after a trial has ended. The rhythmic difference between coherence and streaming (i.e., gallop vs. metronome) is especially useful as a cue to participants for which organization they are hearing. Dotted lines represent the tones that are perceived as connected to each other when perceiving coherence and streaming.

frequency difference is large enough and the repetition rate is sufficiently fast, listeners often report hearing two streams of tones, each in a metronome-like rhythm (i.e., A-A-A-A- and B-B-). The difference in perception of rhythm in the two organizations is particularly useful for experiments measuring subjective reports because participants can use rhythm as a cue to indicate whether they perceive coherence or streaming. As shown in Figure 2, it is also possible to use more objective measures of streaming by introducing patterns that can be detected only when perceiving coherence (e.g., relative timing of adjacent A and B tones; see Figure 2A) or by inserting deviant events that can be detected only when perceiving streaming (e.g., a deviant tone frequency in an otherwise monotonic stream with randomly varying frequency tones in another stream; see Figure 2B).

Although frequency separation and stimulation rate are important determinants in perceiving coherence and streaming, Van Noorden (1975) showed that there are a large number of combinations of frequency separation and stimulation rate that can be heard as either coherent or streaming. It is of more noted importance that he showed that for such ambiguous stimuli it is possible to intentionally bias one's perception toward coherence or streaming. Such findings, as well as subsequent ones, led Bregman (1990) to propose two general types of mechanisms for segregation, primary and schema-based processes. *Primary* mechanisms are those that are largely stimulus driven and can occur outside the

focus of attention. *Schema-based* mechanisms are those that require attention and/or experience-based knowledge (e.g., using knowledge of one's own language to identify speech from background noise).

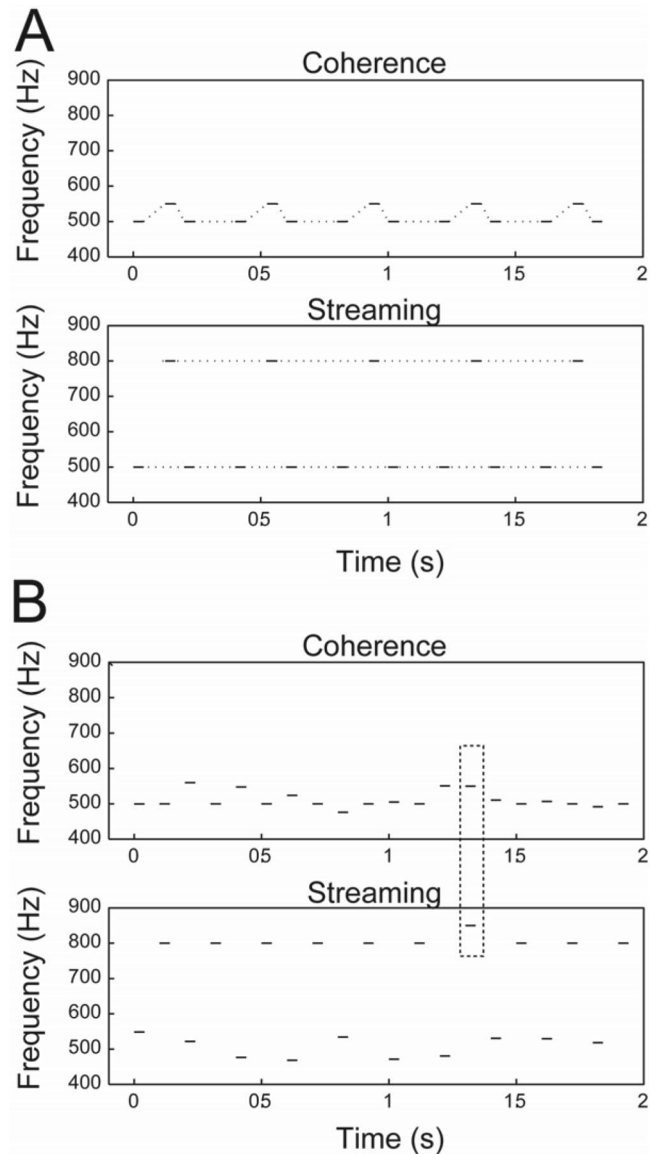


Figure 2. Similar tone patterns as those in Figure 1 can be used for objectively measuring perception of streaming. **A:** An ABA- pattern composed of low tones (A), high tones (B), and silences (-) is shown that is identical to the pattern in Figure 1, except the B tone is shifted in time by a constant amount relative to the A tones. This disrupts the galloping rhythm during coherence (top panel) but does not disrupt the metronome rhythms during streaming (bottom panel). If participants perceive the disrupted rhythm, they are therefore likely to have perceived coherence. **B:** An ABAB pattern in which the A tones vary at random in frequency (within their low-frequency range), whereas the B tones have a constant frequency except for containing one deviant high tone. This deviant tone stands out only if there is a large frequency separation between the A and B tones (bottom panel) but not if there is a small frequency separation between the A and B tones (top panel). If participants perceive the deviant tone, they are therefore likely to have perceived streaming.

The simplicity of the auditory stream segregation paradigm, the ease with which participants can make subjective reports of streaming based on rhythm, the possibility of having the A and B tones differ on one of many acoustic dimensions in addition to various contextual and attentional manipulations, and the paradigm's similarity to more complex situations such as a cocktail party have made it popular for studying auditory scene analysis (for discussion of other paradigms, including those for studying segregation of concurrent sounds, see Alain, in press; Bregman, 1990; Carlyon, 2004; Darwin, 1997). Numerous behavioral studies of stream segregation have thus been published in the past few decades, and their results have told us much about the determinants and mechanisms of streaming (for a detailed review, see Moore & Gockel, 2002). However, a number of important questions and controversies have emerged over the last several years. These include the extent to which attention is necessary for streaming to occur; the extent to which streaming involves processes in the peripheral auditory system versus the central auditory system; and the extent to which streaming involves processes in auditory-specific brain areas versus more modality-general brain areas responsible for perceptual organization or attention. These issues remain the focus of behavioral studies of streaming up to the present.

In addition to traditional behavioral methods, cognitive neuroscience techniques for measuring neurophysiological processing in humans (Posner & DiGirolamo, 2000) and the advancement of knowledge in auditory anatomy and physiology (Kaas & Hackett, 2000) have made possible a deeper understanding of mechanisms underlying auditory stream segregation. Such efforts are beginning to tell us about the various stages of processing involved in streaming and about the anatomical sites important for streaming. Neurophysiological studies are also helping to address issues that traditionally were investigated with behavioral techniques, such as the extent to which attention is necessary for streaming and the nature of the representations underlying streaming. The purpose of this article is to review these new contributions and to reveal an emerging picture of the neural mechanisms required for streaming.

In this article, we will first provide an overview of the anatomical and physiological characteristics of the peripheral and central auditory systems while briefly explaining the relevant neurophysiological techniques for measuring brain activity in humans and nonhuman animals. Next, we will review neurophysiological studies of auditory stream segregation, with particular focus on demonstrating (a) how these studies have contributed to resolving existing controversies and generating new research questions, (b) how they have identified anatomical sites and neurophysiological processes particularly important for streaming, and (c) what future studies using neurophysiological techniques might be helpful in further advancing knowledge on auditory stream segregation.

This review will explore two general issues regarding the mechanisms of auditory stream segregation that have been addressed by recent neurophysiological studies. The first issue is the stages of processing involved in streaming, in particular to what extent streaming relies on peripheral and central mechanisms as well as what neural processes and brain areas are likely to play a role in streaming. The second issue is the role of high-level factors such as attention in streaming. Although the focus of this review is on neural processes, we will also review the relevant behavioral and computational literature with a focus on what these studies have

revealed about mechanisms of streaming. The neurophysiological studies will then be used to show how they have addressed issues that have not been resolved by using purely behavioral measures. Thus, although we attempt to comprehensively cover neural studies of streaming, we are more selective in our review of behavioral and computational studies, focusing on those that inform neural mechanisms of streaming (also see a recent review of studies of streaming focusing on auditory cortex by Micheyl et al., in press).

Organization of the Peripheral and Central Auditory Systems

Understanding how external sound stimuli are processed in peripheral and subcortical structures is crucial for understanding the nature of representations at higher levels (for an introduction to auditory anatomy see Møller, 2006). Sound waves reaching the ears from the environment are first filtered and amplified by the outer ear and middle ear before arriving at the inner ear. Sound pressure waves modulate the ossicles of the middle ear, which in turn generates sound waves in the cochlear fluids of the inner ear. In the cochlea, the sound waves travel along the basilar membrane in a frequency-specific manner, with displacement in response to higher frequencies reaching its maximum on the stiffer basal portion of the membrane (near where the cochlea connects to the middle ear) and displacement in response to lower frequencies reaching its maximum on the more flexible apical portion of the membrane (farthest away from where the cochlea connects to the middle ear). This initial tonotopic representation (spatial mapping of sound frequency) is the basis of many aspects of auditory coding. The local displacement of the basilar membrane results in hair cells at this particular *place* (i.e., portion of the basilar membrane) releasing neurotransmitters that are received by auditory nerve fibers. These auditory nerve fibers receive the neurotransmitters and transmit action potentials to the first stage of central processing, the cochlear nucleus. From the cochlear nucleus, information is sent to a number of other subcortical areas prior to reaching the auditory cortex. Subcortical auditory nuclei are characterized by the preservation of tonotopic mapping, in addition to integrating information across the two ears and across different frequency regions.

A deeper understanding of how auditory cortex is organized has emerged over the last decade. As shown in Figure 3, three distinct auditory cortical regions have been identified: (a) a core region consisting of three primary auditory areas in the medial portion of the superior temporal plane; (b) a belt region consisting of eight secondary areas surrounding the core; and (c) a parabelt region consisting of two secondary auditory areas lateral to the lateral belt areas (Kaas & Hackett, 2000; Rauschecker, 1998; Semple & Scott, 2003). The three core areas (caudal, rostral, and rostrotemporal) in monkeys show characteristics of primary sensory cortex, including selectivity for simple stimulus features such as pure-tone frequency. Each of the core areas receives separate inputs from the auditory thalamic nucleus (medial geniculate body), demonstrating parallel processing of acoustic information at the earliest stage of the auditory cortex. The belt areas in turn receive input from core areas (in addition to the auditory thalamus and modality nonspecific thalamic nuclei) and project to the parabelt areas (in addition to multiple frontal lobe areas). Like the belt areas, the parabelt areas project to the frontal lobe and modality nonspecific temporal

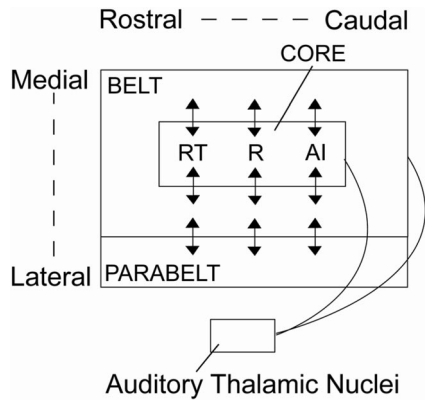


Figure 3. Schematic representation of the organization of old-world monkey auditory cortex in the left superior temporal lobe. Three core areas that receive input from auditory thalamic nuclei (a caudal area, AI; a more rostral area, R; and a rostromedial area, RT) constitute primary auditory cortex. Surrounding the core region is the belt region, which contains a number of secondary auditory cortical areas that receive input from the core region and auditory thalamic nuclei. Lateral to the lateral belt region is the *parabelt* region, which contains additional auditory cortical areas that receive input from the belt areas. Areas in both the belt and parabelt connect to additional cortical areas that are not depicted. Two-way arrows indicate the presence of feed-forward and feedback connections. Humans are likely to have a similar organization of auditory cortex as shown here for old-world monkeys. Based on a figure from “Subdivisions of Auditory Cortex and Processing Streams in Primates,” by J. H. Kaas and T. A. Hackett, 2000, *Proceedings of the National Academy of Sciences of the United States of America*, 97, pp. 11793–11799. Copyright 2000 by the National Academy of Sciences, USA.

regions. In addition to feed-forward connections, feedback connections are also present. Thus, the organization of auditory cortex in monkeys is highly complex, with multiple processing streams and numerous feed-forward and feedback connections within auditory cortex and between auditory cortex as well as other cortical and subcortical brain regions.

Single-cell neurophysiological studies in monkeys have additionally revealed basic aspects of auditory coding that are likely to be present in human auditory cortex. One of the best studied characteristics is the preservation of tonotopy at each stage of processing from the inner ear to the core areas and some belt areas of auditory cortex (Kaas & Hackett, 2000; Merzenich & Brugge, 1973). In the caudal area of the monkey core region, neurons that respond best to higher frequencies are located more medially than are neurons that respond to lower frequencies. In the rostral core area, which is coextensive with the caudal core area, this organization is mirror reversed such that higher frequencies are represented more laterally. Outside of the core, cells typically respond to more complex features than pure-tone frequency.

Anatomical studies in humans suggest a similar organization to that of old-world monkeys, with distinct core, belt, and parabelt regions that can be distinguished cytoarchitecturally (Rademacher et al., 2001; Rivier & Clarke, 1997; Sweet, Dorph-Petersen, & Lewis, 2005). In humans, the core area is most often found in the caudomedial portion of Heschl’s gyrus on the superior temporal plane (Rademacher et al., 2001). Belt and parabelt regions in humans are found on more rostromedial portions of Heschl’s gyrus,

in Heschl’s sulcus, and in the planum temporale (Sweet et al., 2005).

Studies of auditory cortical activity in humans also show similarities to sensory coding principles found in monkeys. For example, microelectrode recording during surgery in 1 epileptic patient demonstrated frequency specificity and tonotopic organization in Heschl’s gyrus, with higher frequencies represented more caudomedially (Howard et al., 1996). A study using functional magnetic resonance imaging (fMRI), a technique that measures hemodynamic changes in specific brain regions that are thought to correlate with neural activity (Logothetis & Wandell, 2004), provided evidence for two tonotopic areas in humans with mirror symmetry, similar to that found in the core region of monkeys (Formisano et al., 2003). The similarity of the monkey and human auditory systems thus underscores the validity of using monkeys as models for human auditory stream segregation, at least for processes occurring at levels up to primary and secondary auditory cortex.

The timing of peaks revealed by intracranial recordings match the components recorded with noninvasive techniques such as electroencephalography (EEG) and magnetoencephalography (MEG; Godey, Schwartz, de Graaf, Chauvel, & Liegeois-Chauvel, 2001), the primary neurophysiological techniques that have been used for studying streaming in humans. EEG and MEG recordings detect neuroelectric and neuromagnetic activity, respectively, with a temporal resolution in the milliseconds range from sensors placed on (EEG) or near (MEG) the scalp. This activity arises from intracranial sources, the locations of which can be determined by using techniques that work backward from the spatial pattern of activity across the scalp at a given point in time (e.g., Picton et al., 1999; Scherg, Vajsar, & Picton, 1989). It is important to note that the presence and characteristics of auditory event-related potentials (ERPs; obtained by averaging EEG segments time locked to stimulus onsets) and event-related fields (similar to ERPs but obtained with MEG rather than EEG, which detects current sources that are tangentially oriented relative to the scalp) to stimulus onsets are particularly well understood. They consist of short-latency (0–10 ms after sound onset), middle-latency (10–50 ms), and long-latency (50–300 ms) responses thought to reflect activity from the brainstem nuclei, primary auditory cortex (e.g., in Heschl’s gyrus), and secondary auditory cortex (e.g., in planum temporale), respectively (Näätänen & Picton, 1987; Picton et al., 1999).

The long-latency responses include the P1–N1–P2 complex (a series of positive, negative, and positive peaks occurring around 50, 100, and 170 ms) that is elicited to sound onsets or offsets and to abrupt changes in sustained sounds. Another long-latency response, called the mismatch negativity (MMN), occurs when a series of frequent standard sounds is followed by a rare deviant sound that differs from the standards in one or more acoustic dimension such as frequency, intensity, duration, or spatial location (for reviews, see Näätänen & Winkler, 1999; Picton, Alain, Otten, Ritter, & Achim, 2000). MMN can also occur when a repeating pattern of tones is followed by a sound that disrupts the pattern (e.g., Alain, Achim, & Woods, 1999; Trainor, McDonald, & Alain, 2002). These long-latency responses have been especially useful for studies of auditory stream segregation, reviewed below.

Processing Stages in Streaming

Peripheral Channeling Versus Central Processing

Acoustic cues to streaming. One way to indirectly identify stages of processing that are important for stream segregation by using behavioral measures is to examine the extent to which different acoustic cues lead to streaming. An influential theory known as the peripheral channeling hypothesis posits that streaming is primarily based on stimulus processing occurring in the auditory periphery (Beauvois & Meddis, 1996; Hartmann & Johnson, 1991; Van Noorden, 1975). According to the peripheral channeling hypothesis, streaming occurs most strongly when there is minimal overlap in the excitation patterns of different sounds in peripheral auditory structures (i.e., the cochlea and the auditory nerve), providing higher levels of the nervous system with clear evidence for clearly distinguishable sound sources. As a test of this hypothesis, Hartmann and Johnson (1991) compared the effectiveness of various acoustic dimensions in the promotion of streaming. The task was to identify pairs of familiar melodies that were interleaved with each other. The notes of the two melodies could differ in one of several characteristics. The most effective cues for segregation in this study were ear of stimulation, frequency range of the melody, and spectrum (i.e., pure tone vs. a harmonic complex), all cues that are based on peripheral coding (i.e., differential stimulation of the right and left ears or tonotopy). Another cue that has an effect on streaming is overall intensity level, with higher overall intensity leading to a decrease in streaming judgments for alternating L and H tones (M. M. Rose & Moore, 2000). This might be due to the broadening of peripheral auditory filters with high intensity sounds, thus leading to more overlap between the excitation patterns of L and H tones (J. E. Rose, Hind, Anderson, & Brugge, 1971). Although these findings are consistent with the peripheral channeling hypothesis, one should keep in mind that just because a cue is extracted very early in the periphery does not mean that there is no further processing at higher levels. For example, pure-tone frequency is not only coded in the inner ear but also in primary auditory cortex (Kaas & Hackett, 2000; Merzenich & Brugge, 1973). Coding of simple cues at higher levels of the auditory system is important because this likely makes it possible to integrate such information with more complex, centrally computed acoustic cues, in addition to information from other modalities.

As a result of the peripheral channeling hypothesis, many studies have further examined the extent to which cues based on central coding can lead to stream segregation (for a detailed review, see Moore & Gockel, 2002). For example, studies have shown that perception of streaming can occur when two sounds differ in amplitude modulation rate (Grimault, Bacon, & Micheyl, 2002), timbre (Cusack & Roberts, 2000; Roberts, Glasberg, & Moore, 2002; Singh & Bregman, 1997), phase (Roberts et al., 2002; but see Stainsby, Moore, Medland, & Glasberg, 2004), and bandwidth (Cusack & Roberts, 1999), in the absence of peripherally based cues. Streaming can also occur when the pitch of two sounds is computed solely on the basis of binaural information, suggesting central integration (Akeroyd, Carlyon, & Deeks, 2005). Another pitch-related cue to segregation that requires processing at central sites is periodicity pitch, the pitch given by the repetition rate of a complex waveform, which corresponds to the fundamen-

tal of the harmonic series into which the waveform can be analyzed. In tones evoking a periodicity pitch, it is still possible to hear the pitch when all low (resolved) harmonics are missing, including the fundamental, hence the term *missing fundamental* used to refer to the perception of pitch in such tones. Periodicity pitch in the absence of spectral cues is based on the periodic acoustic waveform generated by the high (unresolved) harmonics. Vliegen and Oxenham (1999) tested whether streaming could occur from periodicity pitch by presenting ABA- patterns, with alternating tones having different fundamental frequencies but with the harmonics only from 2,000 to 8,000 Hz present. Participants were encouraged to hear the patterns as streaming (cf. Van Noorden, 1975). Streaming with these tones was no different than when the complex tones had low, resolved harmonics (500–2,000 Hz) present or when they used pure tones. These authors subsequently showed that when participants were not biased to hear two streams, they also showed streaming for unresolved harmonic complexes, but the percept was not as strong as with pure tones or resolved harmonic complexes (Vliegen, Moore, & Oxenham, 1999; also see Grimault, Micheyl, Carlyon, Arthaud, & Collet, 2000).

Effects of peripheral impairment on streaming. Another way to evaluate the role of peripheral processing on streaming is by testing people with hearing loss that accompanies normal aging (i.e., presbycusis), people with clinical hearing impairment, and people with cochlear implants. Studies have provided evidence for normal stream segregation of tones differing in pure-tone frequency in healthy aging by using both subjective (Snyder & Alain, 2007) and objective (Alain, Ogawa, & Woods, 1996; Trainor & Trehub, 1989) behavioral measures of stream segregation, in addition to ERP measures (Snyder & Alain, 2007), suggesting that mild peripheral hearing loss typical of aging does not impair streaming. M. M. Rose and Moore (1997) showed that some (but not all) hearing-impaired listeners required greater frequency separation than did normal-hearing listeners for stream segregation to occur. Similar individual differences are present in people with cochlear implants, some of whom showed impairment in streaming and some of whom did not (Hong & Turner, 2006; also see Chatterjee, Sarampalis, & Oba, 2006; Cooper & Roberts, 2007). Furthermore, the impairment in streaming among some cochlear implant users could not be explained by differences in the frequency resolution of the cochlear implant device. However, recent data suggest that caution should be taken in interpreting studies of cochlear implant users because these studies do not show evidence of some aspects of streaming such as rate dependence and perceptual reversals (Cooper & Roberts, 2007). Thus, studies that report perception of streaming in cochlear implant users may instead be measuring frequency discrimination or some other frequency-based perceptual phenomenon other than streaming.

In a study using harmonically complex tones, young adults perceived more streaming than did older adults but only for tones with peripherally resolved harmonics (Grimault, Micheyl, Carlyon, Arthaud, & Collet, 2001), consistent with peripheral hearing deficits in the older adults. When the older adults were separated into two groups, however, those with hearing impairment perceived more streaming than did those without hearing impairment for the resolved tones, but only when the stimulus level was 30 dB.

This result should be evaluated cautiously, though, because there were only 5 participants in the older normal-hearing group. For tones with peripherally unresolved harmonics requiring central processing, robust streaming was observed for all groups of participants.

The above studies thus show inconsistencies in terms of whether peripheral changes due to normal aging, hearing loss, or having a cochlear implant have a detrimental impact on streaming. Note that changes in criteria for subjective judgments of streaming as a result of mild or severe hearing loss could account for some of the null findings. However, this is less likely to account for studies using objective measures of streaming that do not find differences due to aging (Alain et al., 1996; Trainor & Trehub, 1989). Consistent with studies showing that differences in centrally computed sound features can lead to streaming, even severe impairment in peripheral auditory processing does not always impair streaming, even for cues that are based on peripheral coding such as pure-tone frequency. The inability of the peripheral channeling hypothesis to completely account for cues that lead to streaming and to predict the effects of peripheral hearing impairment on streaming highlights the importance of central mechanisms in auditory stream segregation. Thus, a strict interpretation of the peripheral channeling hypothesis, that all aspects of streaming occur in the periphery, is untenable.

Using ERPs as Neurophysiological Landmarks

One way to use neurophysiological activity to inform how the nervous system carries out streaming is to use ERPs as temporal landmarks to identify distinct stages of processing. Early ERP studies focused on the interaction between streaming and selective attention (Alain, Achim, & Richer, 1993; Alain & Woods, 1994). In auditory selective attention tasks, participants are usually presented with two streams of sounds, with each stream being defined by its frequency and/or location. In these tasks, participants are asked to focus their attention on one particular stream to detect occasional sounds that slightly differ from the standard sounds defined by the task-relevant streams. Prior research has shown enhanced ERP amplitude for sounds occurring within the attended stream relative to the same sound when attention was directed elsewhere (Hillyard, Hink, Schwent, & Picton, 1973). It is interesting to note that the effects of selective attention on ERPs are larger in situations that promote streaming (for a review, see Alain & Arnott, 2000). More specifically, grouping sounds based on either frequency or location facilitated the processing of task-relevant stimuli and at the same time decreased the interference of task-irrelevant stimuli (Alain & Woods, 1994). Alain and Arnott (2000) proposed an object-based account in which auditory attention is allocated to auditory objects after an initial process that partitions the incoming events into distinct streams according to Gestalt principles.

The notion that auditory streams can be formed outside the focus of attention was examined in subsequent studies using the MMN. Although MMN amplitude can be modulated by attention (e.g., Alain & Izenberg, 2003; Näätänen, Paavilainen, Tiitinen, Jiang, & Alho, 1993; Woldorff, Hillyard, Gallen, Hampson, & Bloom, 1998), it can be recorded during passive listening conditions (no response required) and may therefore

reflect an automatic change detection process. The basic strategy of using MMN as a marker for determining the level of processing at which stream segregation occurs is to construct sequences of tones made of two potentially separable streams, with at least one of the streams containing a series of standard and deviant tones. If the two streams differ sufficiently in frequency range, the auditory system should be able to process the two streams in separate frequency channels. Thus, the channel that is processing the frequency range containing standards and deviants should be able to detect the deviants and generate MMN, without being affected by the tones in the other stream. This approach is supported by evidence that the MMN can index detection of deviants from tones that alternated between the two ears (McKenzie & Barry, 2006; Ritter, Sussman, & Molholm, 2000).

An early study that used MMN to study streaming used alternating H and L tones that were presented while participants read self-selected media (Sussman, Ritter, & Vaughan, 1999). Both the H and the L tones contained tone triplets that ascended in frequency (i.e., L1,L2,L3 and H1,H2,H3). Occasionally, this ascending pattern was disrupted by presenting a single pattern of descending tones (i.e., L3,L2,L1 or H3,H2,H1). The patterns were presented at different stimulation rates: 100-, 250-, or 750-ms stimulus onset asynchrony (SOA) between H and L tones. The stimuli with the shortest SOA caused streaming, according to participants' informal subjective reports following the experiment, whereas the stimuli with the longest SOA did not. The authors did not report whether perception of streaming occurred for the 250-ms SOA. Following the disrupted pattern, MMN occurred to the L-tone and H-tone disruptions in the 250-ms SOA but only to the L-tone disruption in the 100-ms SOA, suggesting that MMN may index streaming. However, the less robust MMN for the 100-ms SOA pattern compared with the 250-ms SOA pattern suggests that this is an indirect reflection of processes related to streaming. Also problematic is the lack of MMN in a control condition presenting just the L tones of the 750-ms condition. Thus, MMN does not always occur when a single stream is presented without a competing stream, making it difficult to determine whether the lack of an MMN in a particular condition is due to lack of streaming or some other factor.

Another MMN study provided evidence that streaming occurs before temporal integration (Yabe et al., 2001). Alternating H and L tones were presented with an SOA of 125 ms to participants, who were more likely to report hearing the pattern as two separate streams when the frequency separation was large (tones of 3,000 Hz and 500 Hz) than when the frequency separation was small (tones of 3,000 Hz and 2,800 Hz). Occasionally, one of the H tones was omitted from the pattern. When the frequency separation between H and L tones was small, a clear MMN to the omitted tone occurred because the time between adjacent H and L tones was less than the 150-ms temporal integration window known to be involved in MMN elicitation to tone omissions (Tervaniemi, Saarinen, Paavilainen, Danilova, & Näätänen, 1994; Yabe et al., 1998). When the frequency separation was large, however, no MMN occurred in response to the omitted tone, presumably because the H and L tones were processed in separate streams and because the SOA between adjacent tones in the H stream was 250 ms. These results suggested that a form of temporal inte-

gration occurs after frequency-based segregation but before MMN generation (also see Sussman, 2005).

Winkler, Takegata, and Sussman (2005) investigated the relationship between temporal integration, frequency-based segregation, and perception of streaming. They presented participants with an ABA- pattern that could be heard as one stream or two streams. Occasionally the B tone was omitted and the authors reported two different negative difference waves in response to the omission, an early wave and a later wave. The early difference wave occurred whether participants perceived one or two streams or none, whereas the later difference wave occurred only when participants perceived one stream. These data suggest the presence of an early stage of temporal integration, which reflects stimulus configuration, and a later stage of temporal integration, which is influenced by whether the ABA- pattern is perceptually integrated into a single stream.

Sussman, Horváth, Winkler, and Orr (2007) used the MMN to provide evidence that the MMN reflects buildup of streaming (Anstis & Saida, 1985; Bregman, 1978) by comparing brain responses to intensity deviants in a sequence of H tones that were interleaved with L tones that varied in intensity at random. When the frequency separation between the L and H tones was large, an MMN to the intensity deviant occurred late in the sequence after buildup was likely to have occurred. No MMN occurred when the frequency separation was small, even late in the sequence. Thus the MMN seems to reflect streaming buildup because it occurred only late in the sequence and when the frequency separation was large. Because participants were ignoring the sound stimuli by performing a difficult noise-change detection task, the authors further argued that buildup could occur in the absence of attention (for more details, see *The Role of Attention*).

From studies using MMN as an index of streaming, it is not clear how the MMN and streaming are interrelated. In other words, the MMN could simply reflect segregation of tones into different frequency channels at a stage of processing preceding the MMN without indexing perception of streaming per se (but see Winkler et al., 2005). For example, one study showed that MMN could occur in response to low- and high-frequency deviants in tones that alternated in frequency with a 500-ms SOA, despite the fact that participants did not perceive streaming (Shinozaki et al., 2000). This is in contrast to the study by Sussman et al. (1999) who showed MMN only in sequences that were presented rapidly enough to cause streaming. In another study, MMN to a frequency deviant occurred when tones were alternated with a 1,000-ms SOA and a one semitone difference (corresponding to adjacent white and black keys on a piano) between the alternating tones, a pattern that was highly unlikely to result in perception of streaming (Alain & Woods, 1994). Thus, it remains a possibility that the MMN may be a poor indicator of whether participants perceive streaming. Another possibility is that MMN correlates with streaming under conditions in which only the unsegregated pattern is too complex for the deviant to be picked up, whereas the simpler within-stream patterns permit the deviant to be detected. Such an interpretation would suggest that MMN is not a direct correlate of processes related to streaming but rather that MMN is elicited in many, but not all, situations that promote streaming.

Another point of caution is necessary when using the MMN (and other ERPs) to make conclusions about stages of processing. Although it is known that the MMN arises primarily from neural generators in the superior temporal plane, this does not necessarily mean that the MMN occurs at a later anatomical stage than do those processes that elicit the MMN. For example, it is possible that some processes involved in streaming occur at secondary auditory areas or higher level multimodal brain regions and provide feedback signals to the MMN generators. Similarly, the processes that are involved in streaming and those involved in MMN generation could operate in parallel, consistent with the idea described above that some but not all of the conditions that elicit MMN also promote streaming. Although such a view of auditory processing may seem overly complex, there is strong evidence from anatomical studies of the auditory system for parallel and feedback processing in widely distributed networks of auditory areas (Kaas & Hackett, 2000; Winer, Miller, Lee, & Schreiner, 2005). Thus, an important goal for neurophysiological studies of streaming is to more directly identify neural processes and their brain substrates involved in specific aspects of streaming.

Neural Processes and Brain Areas Involved in Streaming

Measuring sensory-evoked activity that occurs while presenting tones with differing acoustic attributes might allow for a more direct link between perception of streaming and neural activity involved in stream segregation. Such an approach was recently taken by two independent research groups, who recorded auditory neuromagnetic and neuroelectric responses (respectively, Gutschalk, Micheyl, Melcher, Rupp, Scherg, & Oxenham, 2005; Snyder, Alain, & Picton, 2006). Both of these studies varied the frequency separation between A and B tones in a repeating ABA- pattern. The sensory-evoked response time locked to the B tone increased in amplitude as a function of frequency separation, and this increase was highly correlated with perception of streaming in individual participants (Gutschalk et al., 2005; Snyder et al., 2006). In the ERP domain, a frequency-dependent modulation of the P1, N1, and P2 was maximal at frontocentral electrodes, whereas an N1c modulation occurred only at the right temporal electrodes (Snyder et al., 2006), as shown in Figure 4A (for more details see *The Role of Attention*). Dipole source modeling revealed bilateral tangential sources accounting for P1–N1–P2 (Gutschalk et al., 2005; Snyder et al., 2006) and a right hemisphere radial source accounting for N1c (Snyder et al., 2006) in or near Heschl's gyrus. Gutschalk et al. (2005) also found, for ABA- patterns with the same frequency separation, larger N1 amplitude when participants reported hearing two streams than when they reported hearing a single stream. This suggested that the changes in sensory-evoked response did not solely reflect stimulus-driven activity but was also indicative of perception of streaming per se.

Such increases in long-latency response amplitudes with increasing frequency separation may be interpreted as a release from

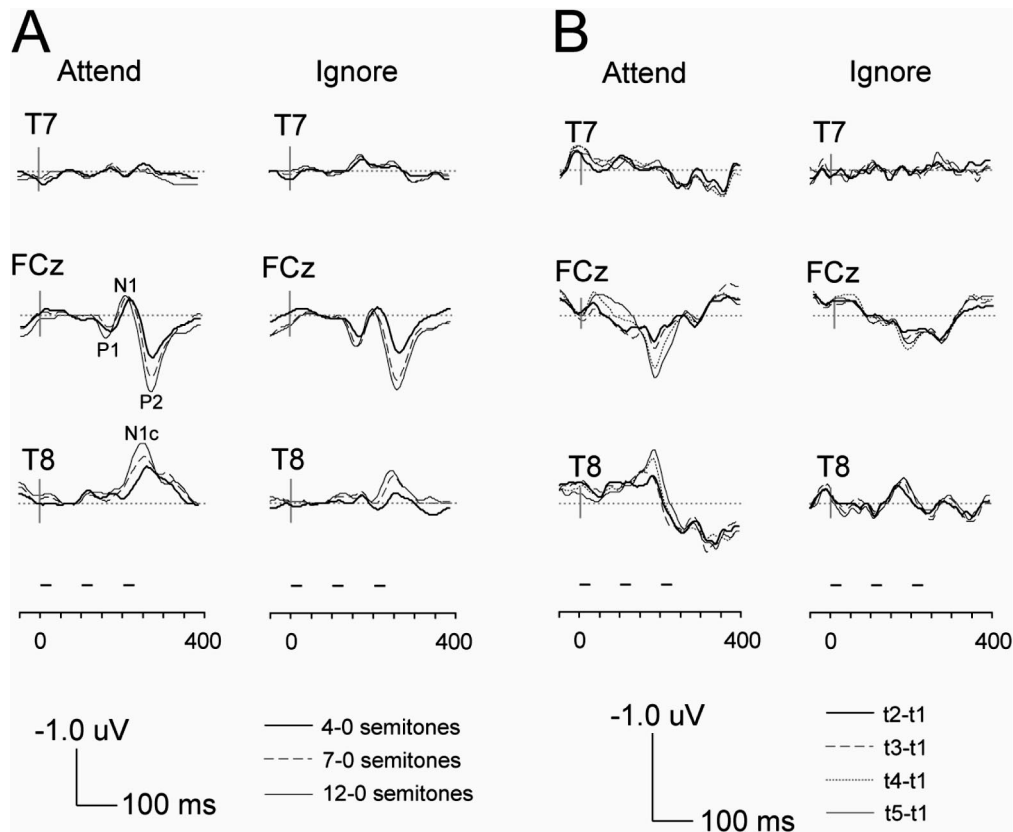


Figure 4. A: Difference waves showing the effect of frequency separation on the P1–N1–P2 and N1c responses. Difference waves are calculated by subtracting ERPs elicited by ABA- patterns with 0 semitone frequency difference from those elicited by 4, 7, and 12 semitone frequency differences when participants were listening and making streaming judgments at the end of the trial (attend condition, left column) and when they were ignoring the stimuli by watching a muted, subtitled movie (ignore condition, right column). Difference wave event-related potentials (ERPs) are shown for electrodes at T7 (left temporal scalp region), FCz (fronto-central midline scalp region), and T8 (right temporal scalp region), showing rightward asymmetry for the N1c response. Horizontal bars above the time scale represent pure tones in the ABA- pattern. B: Difference waves showing the effect of buildup on the ERP for attend (left column) and ignore (right column) conditions. Difference waves are calculated by subtracting ERPs elicited by ABA- patterns in the first 2 s of the trial (t1) from those elicited by the next four 2-s time bins (t2, t3, t4, t5). Data are from “Effects of Attention on Neuroelectric Correlates of Auditory Stream Segregation” by J. S. Snyder, C. Alain, and T. W. Picton, 2006, *Journal of Cognitive Neuroscience*, 18, 1–13. Copyright 2006 by the Massachusetts Institute of Technology.

neural suppression.¹ Specifically, when the same tone is repeated at short SOAs, the response magnitude to individual tones decreases (Carver, Fuchs, Jantzen, & Kelso, 2002; Davis, Mast, Yoshie, & Zerlin, 1966; Hari, Kaila, Katila, Tuomisto, & Varpula, 1982; Lu, Williamson, & Kaufman, 1992; Snyder & Large, 2004). This may indicate adaptation, habituation, inhibition, or short-term synaptic depression because of repeated stimulation. When an acoustic change is introduced, however, response magnitude often recovers (Butler, 1968; S. J. Jones, Longe, & Vaz Pato, 1998; Martin & Boothroyd, 2000; Näätänen et al., 1988; Picton, Woods, & Proulx, 1978), possibly indicating the activation of a new population of neurons responding to the new sound. This recovery from response suppression after acoustic change may therefore be consistent with the idea that stream segregation depends on activation of distinct populations of neurons in auditory cortical regions.

By using relatively short (10.8-s) tone sequences, Snyder et al. (2006) were also able to examine neural activity linked to the buildup of streaming observed in behavioral studies (Anstis & Saida, 1985; Bregman, 1978). Sequences were divided into five different time bins (each 2 s long), and the ERPs were averaged separately for each time range. A positive enhancement of activity occurred at frontocentral scalp locations that peaked about 200 ms after the beginning of each repetition of the ABA- pattern and increased over the course of the sequence, paralleling the perceptual buildup of streaming that occurs over several seconds (see Figure 4B; for more details, see *The Role of Attention*). As with the

¹ We use the term *suppression* to mean any reduction in response, without reference to a specific neurophysiological mechanism that would result in response reduction.

frequency separation-related activity, this buildup activity was mostly accounted for by bilateral tangential sources and a right radial source in or near Heschl's gyrus. The specific physiological process generating this time-dependent modulation is currently not well understood. However, one possibility is that it may result from an attention-dependent negativity called the negative difference wave (Hansen & Hillyard, 1980) being present at the beginning of the trial and becoming smaller during the course of the trial. Such an interpretation would depend on greater attention being allocated to the beginning of the trial compared with later in the trial. However, this interpretation may not be valid because negative difference increased over time, rather than decreased, when participants were selectively attending to one of two series of interleaved tones (Hansen & Hillyard, 1988).

In addition to human EEG and MEG studies of auditory streaming, fMRI has provided additional information about the neural structures involved in stream segregation. One study showed that posterior regions of left auditory cortex were modulated by listening to alternating organ and trumpet tones when compared with a single stream of either organ or trumpet tones presented at the same rate, while detecting targets in one of the two streams (Deike, Gaschler-Markefski, Brechmann, & Scheich, 2004). This is consistent with EEG and MEG studies discussed above showing frequency-related modulations in auditory cortex, although the right temporal lobe rather than the left appeared to be dominant (Snyder et al., 2006). An important question is the extent to which the activation reported by Deike et al. (2004) was the result of spectral encoding versus a more categorical representation of the two instruments.

Another fMRI study used pure-tone stimuli and showed differential activity in the intraparietal sulcus, depending on whether participants heard one or two streams for ABA- patterns with the same frequency separation (Cusack, 2005). Activations in such nonauditory regions might index higher level processes such as object formation or auditory attention to objects. Surprisingly, no frequency separation-dependent activation was observed in auditory cortex.

This is in contrast to a more recent study showing fMRI signal increases in primary auditory cortex and planum temporale with increasing frequency separation in a pure-tone ABAB pattern (Wilson, Melcher, Micheyl, Gutschalk, & Oxenham, 2007). Similar findings occurred in a study measuring fMRI activity while participants listened to ABA- patterns with complex tones differing only in fundamental frequency but not spectral distribution (Gutschalk, Melcher, Micheyl, Wilson, & Oxenham, 2006), with fundamental frequency separation-related increases of activity mainly in secondary auditory cortex (i.e., outside of Heschl's gyrus). Gutschalk et al. (2006) also reported an increased P1 response recorded with MEG that paralleled the separation in fundamental frequency between the A and B tones, similar to what was found in previous studies using pure tones (Gutschalk et al., 2005; Snyder et al., 2006). This suggests that segregation processes reflected in auditory cortical activity are not limited to differences in activation along a tonotopic representation, but they might reflect a more general mechanism for segregation of neural activity between populations coding for stimuli that differ in a number of properties. This is consistent with previous ERP studies showing recovery of the N1 following changes in a wide variety of dimensions, such as frequency, spectrum, location, and intensity

(Butler, 1968; Hung, Jones, & Vaz Pato, 2001; S. J. Jones et al., 1998; Martin & Boothroyd, 2000; Näätänen et al., 1988).

Recording from single or multiple neurons in nonhuman animals has provided additional information regarding the mechanisms that underlie streaming. The validity of studying neural mechanisms of streaming in nonhuman animals is underscored not only by similarities between the anatomy and physiology of different species' auditory systems but also by perceptual demonstrations of stream segregation in a wide variety of species, including fish, birds, monkeys, and bats (Fay, 1998, 2000; Hulse, MacDougall-Shackleton, & Wisniewski, 1997; Izumi, 2002; Moss & Surlykke, 2001; for a review, see Hulse, 2002). A study in European starlings provided particularly strong evidence for close correspondence to human perception of streaming (MacDougall-Shackleton, Hulse, Gentner, & White, 1998). The starlings were first trained to peck one key when listening to a constant frequency AAA- pattern in a galloping rhythm and to peck a different key when listening to a single stream of tones either at the tempo of the A tones (i.e., A-A-) or at the tempo of the B tones (i.e., B—). When presented with ABA- tone patterns, the birds were more likely to press the key corresponding to the streaming patterns when frequency separation between the A and B tones increased. This mirrors human perceptual reports of the change in rhythm that accompanies streaming. Despite the evidence that several species of nonhuman animals perceive streaming similarly to humans, the studies reviewed below that recorded neural activity in nonhuman animals should be evaluated cautiously because these studies did not collect behavioral responses from the same animals to compare with their neural activity.

Fishman and colleagues recorded multiunit activity and current source densities from primary auditory cortex neurons of awake macaque monkeys (Fishman, Arezzo, & Steinschneider, 2004; Fishman, Reser, Arezzo, & Steinschneider, 2001). During the recordings, they presented tones that alternated in frequency in an ABAB pattern, with different frequency separations, presentation rates, and tone durations. It should be noted that in the study by Fishman et al. (2001), each sequence of tones was fixed at 490 ms such that for the slowest presentation (5 Hz) rate only three tones were presented, making it unlikely that perception of streaming occurred. However, Fishman et al. (2004) used longer sequences and obtained similar results. These studies recorded from neurons that were tuned to frequencies corresponding to the A tone. Thus, any difference in the responses to the A and B tones would suggest that distinct neural populations were processing the two tones. Increasing frequency separation, increasing presentation rate, and using longer tone durations caused suppression of responses to the B tones relative to the A tones, paralleling human perception of streaming (Moore & Gockel, 2002). Responses to the A tones also decreased with increasing presentation rate, although less so than responses to the B tones. The authors proposed that physiological forward masking could account for the suppression of B-tone responses.

A study that recorded intracortical ERPs and single-unit activity from the primary auditory cortex of awake mustached bats also found increased responses to A tones relative to B tones (in neurons tuned to A tones) with increasing frequency separation and presentation rate (Kanwal, Medvedev, & Micheyl, 2003). These authors proposed a model that incorporated recurrent inhibition and lateral inhibition to simulate the ability of single neu-

rons to respond primarily to the A tones and to suppress responses to the B tones.

A recent study extended findings of neural stream segregation to insects. Schul and Sheridan (2006) recorded from neurons in katydid that responded to a simulated bat echolocation call. The bat call was presented with a frequency of 40 kHz and a repetition rate of 7 Hz, interleaved with a train of simulated katydid sounds with a frequency of 15 kHz and a repetition rate of 140 Hz, after being adapted with the katydid songs alone. When the bat and katydid sounds both were presented with a frequency of 15 kHz, no additional response to the bat call occurred after adaptation to the katydid song, demonstrating that the change in temporal pattern of pulses when inserting the bat sound was not sufficient to elicit a response. These studies suggest that frequency-based segregation occurs in the auditory systems of a wide variety of species, including both vertebrates and invertebrates, and suggest an important role of suppression mechanisms in frequency-based segregation.

Although these studies demonstrate the possibility that neural processes in auditory cortex form the basis of stream segregation, it is important to compare perceptual data with neural responses to determine how closely the two are related. To this end, Bee and Klump (2004) compared multiunit responses in the auditory forebrain of songbirds to behavioral data from humans. They found a close correspondence between the neurophysiological data from songbirds and the fission boundary in humans, the threshold below which people can no longer hear two streams even when actively trying to hear streaming (Van Noorden, 1975). This study also tested for evidence of forward suppression. The authors found greater suppressive effects of the first A tone on the B tone compared with the effects of the B tone on the second A tone in the ABA- pattern, confirming a role for suppression of tones outside the characteristic frequency range of single neurons. Such suppression likely serves to enhance the spatial segregation of neural responses to A and B tones in primary auditory cortex.

Micheyl, Tian, Carlyon, and Rauschecker (2005) further explored the correspondence between single-unit recordings of primary auditory cortical neurons that were tuned to the A-tone frequency in macaque monkeys and human perception of streaming. On individual trials, responses to A and B tones decreased over multiple repetitions of the ABA- patterns. Furthermore, the decrease in B-tone response over time was more pronounced for larger frequency separations, whereas the decrease in the A-tone response did not depend on frequency separation. A simple response magnitude threshold was applied to the data as a model for how the neural activity might be used by other brain areas to decide whether a particular neuron was detecting one stream or two streams. The output of this simple model showed good correspondence with the time course of perceptual buildup from a group of human participants. This study extends previous research implicating suppression mechanisms in frequency-based segregation by also showing that suppression might play an important role in the buildup process that accompanies streaming. Consistent with these findings, an ERP correlate of buildup was localized to auditory cortex (Snyder et al., 2006), and a computational model of stream segregation required inhibitory time constants typical of cortical processing to successfully model the perceptual buildup of stream segregation (McCabe & Denham, 1997). However, a previous model of streaming that assumed frequency-based segrega-

tion occurring exclusively in the auditory periphery was also able to model buildup processes by implementing neural response adaptation in the auditory nerve (Beauvois & Meddis, 1996). Further computational modeling directly comparing models based on different levels of processing may be necessary to address the extent to which peripheral, subcortical, and cortical processes are involved in streaming.

Collecting neurophysiological data from subcortical auditory nuclei in addition to auditory cortex in nonhuman animals may be important for determining the extent to which frequency-based segregation and buildup processes observed in human and animal data at the level of auditory cortex are manifestations of subcortical processes. Recently, Pressnitzer, Micheyl, Sayles, and Winter (2007) recorded from single ventral cochlear nucleus neurons tuned to A tones in anesthetized guinea pigs while presenting ABA- patterns similar to those used by Micheyl et al. (2005). Similar frequency separation dependence and long adaptation effects were observed as in Micheyl et al. (2005), suggesting that these aspects of streaming are present at the earliest stages of central auditory processing, even in anesthetized animals. An important question arising from this finding is whether buildup is an intrinsic aspect of the computations required for cue-based segregation. Such a hypothesis would predict that the brain areas responsible for streaming buildup depend on (and are the same as) those that compute any particular cue (cf. Micheyl et al., 2005; Snyder et al., 2006). Another possibility to investigate is whether segregation with cues other than pure-tone frequency relies on topographic representations, for which there is some indirect support (Langner, 1997; Schulze & Langner, 1997; but see McAlpine, 2004).

It is unclear what biophysical mechanisms are responsible for the neurophysiological suppression that appears to be related to the buildup of streaming. Given that streaming occurs best for patterns with short interstimulus intervals (Bregman, Ahad, Crum, & O'Reilly, 2000), it is possible that inhibitory gamma-aminobutyric acid (GABA) neurons, which suppress excitatory activity for up to 200 ms following a stimulus (Dutar & Nicoll, 1988; McCormick, 1989), account for streaming-related suppression. This would also be consistent with the importance of GABAergic inhibition in frequency tuning precision (Chen & Jen, 2000; Tan, Zhang, Merzenich, & Schreiner, 2004; Wang, McFadden, Caspary, & Salvi, 2002; Wehr & Zador, 2003). However, this type of suppression cannot account for streaming in patterns with long interstimulus intervals or for the longer temporal processes observed in neural (Micheyl et al., 2005; Snyder et al., 2006) and behavioral (Anstis & Saida, 1985; Beauvois & Meddis, 1997; Bregman, 1978; Snyder, Carter, Lee, Hannon, & Alain, 2007) studies. It is therefore likely that additional suppression mechanisms acting over longer time spans play a role in streaming (Ulanovsky, Las, Farkas, & Nelken, 2004; Ulanovsky, Las, & Nelken, 2003; Wehr & Zador, 2005).

An important limitation to the animal studies discussed above is that behavioral responses were not collected to directly compare with the neural data in the same animals. This warrants a note of caution in interpreting how closely the neural responses in auditory cortex relate to perception of streaming, rather than simply reflect frequency separation-dependent responses. A related issue is the possible role of cortical regions outside of primary auditory cortex in performing various processes important for streaming. Areas

outside of primary auditory cortex might play a role in segregation by using more complex acoustic features (e.g., periodicity pitch, Gutschalk et al., 2006), in addition to supporting processes that are more directly tied to perceptual organization (Cusack, 2005). A final limitation with animal studies is the difficulty in manipulating variables such as attention and schematic knowledge, which would require extensive behavioral training of the animals prior to neurophysiological recordings. If successful, however, recording of behavioral and neural responses in awake nonhuman animals could provide important information about the mechanisms of streaming that may never be gained from human neuroimaging studies. Another potentially fruitful methodology is to record neural activity intracranially from awake humans undergoing neurosurgery (e.g., Howard et al., 1996).

High-Level Factors in Streaming

Development of Primary and Schema-Based Processes

Despite the emerging evidence that streaming reflects a number of different processes, some of which are automatic or primary (i.e., stimulus driven) and some of which are controlled or schema-based (i.e., requiring attention and/or knowledge), the precise role of high-level factors in streaming has remained a topic of intense debate. Support for primary aspects of streaming occurring without much auditory experience or knowledge comes from studies of very young infants. One study found evidence of streaming in 1.5 to 3-month-old infants by using visual fixation (Demany, 1982) and another in newborns by using nonnutritive sucking (McAdams & Bertoni, 1997), both studies showing dishabituation following pattern changes that could be noticed only if streaming had occurred. A study of newborns using an auditory oddball paradigm showed that the ERP difference between standards and deviants interleaved with distracter tones was similar to the ERP difference when the standards and deviants were presented without distracters (Winkler, Kushnerenko, et al., 2003). More complex forms of streaming develop relatively fast, as evidenced by the fact that infants under 1 year old used word knowledge, voice familiarity, and synchronized visual information to segregate speech from a distracting background (Barker & Newman, 2004; Hollich, Newman, & Jusczyk, 2005; Newman, 2005; Newman & Jusczyk, 1996), even when the target and distracter stimuli were presented at equal loudness (Hollich et al., 2005). Thus, developmental studies have provided evidence that primary mechanisms of streaming are present and functioning very early in life; studies of older infants have provided evidence for the rapid development of schema-based mechanisms, supporting the functional importance of streaming mechanisms. Future studies should take advantage of the possibility of noninvasive neurophysiological measurements in infants and older children (Munakata, Casey, & Diamond, 2004) to determine how neural mechanisms of streaming develop early in life.

The Role of Attention

Studies that directly test the effect of attention are necessary to determine the extent to which different aspects of streaming require attention. One study provided evidence that, even under conditions that might be expected to cause streaming with only

primary mechanisms, participants' active listening was required for them to perceive streaming (Carlyon, Cusack, Foxton, & Robertson, 2001). The authors continuously monitored the buildup of streaming by asking listeners to report whether they were hearing one stream or two streams throughout a 21-s ABA- pattern presented to the left ear. On half the trials, listeners were presented with distracting sounds in the right ear, on which they made judgments for the first 10 s of the trial, after which they switched their attention to the ABA- patterns to make streaming judgments. When there was no distracting task, the proportion of time that participants reported two streams showed a normal buildup and reached a plateau at around 10 s. When there was a distracting task, however, streaming was substantially reduced at the point of attention switching. In a second experiment, making amplitude modulation judgments on the ABA- pattern had no such diminishing impact on streaming judgments. Similar findings showed that detecting targets within the ABA- patterns caused less reduction in streaming than did a distracting visual task (Carlyon, Plack, Fantini, & Cusack, 2003). Carlyon et al. (2001) provided converging evidence of the importance of attention to the ABA- patterns for perception of streaming by comparing individuals with unilateral neglect for the left side of space with normal control participants. Compared with control participants, perception of streaming was reduced in neglect patients when stimuli were presented to the left ear but not when stimuli were presented to the right ear.

An independent group of researchers found that adding tones differing in frequency from the ones that participants were attending did not affect temporal irregularity detection within the attended stream (Brochard, Drake, Botte, & McAdams, 1999). However, Alain and colleagues (Alain & Woods, 1993, 1994; Arnott & Alain, 2002) showed that decreasing the frequency separation or spatial separation between different distracters facilitated detection of deviant targets in a designated attended stream, suggesting that perceptual grouping of task-irrelevant stimuli may occur outside the focus of attention. However, the facilitative effect of distracter clustering may have been caused by mutual suppression of the different distracter stimuli. The increased similarity between distracters may have also reduced their saliency, thereby increasing the likelihood of target detection within the task-relevant stream.

A subsequent study directly challenged the conclusion that streaming requires attention by showing that when performing a visual memory task, repeating ABA- patterns that would lead to streaming caused a reduction in interference with the visual task, suggesting that streaming occurred outside the focus of attention (Macken et al., 2003). Although it is possible that the participants in this study attended to the sounds to a small extent, the presence of a difficult visual task makes it unlikely that they attended to the stimuli enough for perceptual buildup to occur as in the study by Carlyon et al. (2001). It is possible that the conflicting results of Macken et al. (2003) and Carlyon et al. (2001) can be reconciled by proposing that the two methods indexed different aspects of streaming. For example, it is possible that Macken et al. (2003) showed that segregation of sounds could occur preattentively, whereas Carlyon et al. (2001) showed that perceptual buildup of streams required attentive processing. Another possibility, which has yet to be ruled out, is that switching attention from one stimulus to another caused a reduction in buildup that occurred outside the focus of attention (Carlyon, 2004; Cusack, Deeks, Aikman, & Carlyon, 2004; Moore & Gockel, 2002). This would be

consistent with resetting of stream biasing that occurs with abrupt changes of the ABA- pattern (Anstis & Saida, 1985; Cusack et al., 2004; Rogers & Bregman, 1998). Thus, behavioral studies on the role of attention in streaming have yet to resolve the issue of whether streaming can occur outside the focus of attention.

A number of studies have used ERPs to determine the role of attention by identifying neural responses that reflect streaming even when participants ignore the stimuli. Recording ERPs has the advantage that distinct stages of processing can be directly identified, some of which may be independent of attention and some of which may be dependent on attention. In contrast, behavioral techniques generally have access only to the output of the many stages of processing that are likely to precede a perceptual report. Using ERP measurement also sidesteps the potential confound discussed above that arises when asking participants to switch their attention to the ABA- patterns after having ignored them for several seconds.

To test the effects of attention on the MMN index of streaming, Sussman and colleagues used the same ascending and descending frequency patterns as in one of their other studies (Sussman et al., 1999), but with an SOA of 500 ms, resulting in an ambiguous pattern that could be perceived as one stream or as two streams (Sussman, Ritter, & Vaughan, 1998). In the first part of the experiment, participants were instructed to ignore the stimuli by reading a book. Next, the same participants were instructed to attend to the H tones and ignore the L tones by pressing a button in response to H-tone deviants. When participants read a book, MMN did not occur to either the L-tone or the H-tone deviants. In contrast, when participants attended to the H tones, an MMN occurred to the L-tone and the H-tone deviants, despite the fact that they were attending to only one of the streams. These results suggested that attention to the H tones facilitated segregation of the L and H tones and allowed the auditory system to detect changes occurring in either of the two streams. This is consistent with participants' informal subjective reports that they perceived streaming while attending to the H tones.

A subsequent study tested the effects of attention on the MMN index of streaming, in addition to comparing the MMN with behavioral responses to make sure that streaming occurred (Winkler, Sussman, et al., 2003). Deviant tones shorter in duration than the standards were presented alone, with distracter tones of various durations in the same frequency range or with distracter tones of various durations in a different frequency range. Participants were equally good at detecting the deviant tones when the standards and deviants were presented alone compared with when they were presented with distracter tones in a different frequency range. However, deviant detection was impossible when the distracter tones were in the same frequency range as the standards and deviants. MMN followed a similar pattern by occurring only when the standards and deviants were presented alone or with distracters in a different frequency range. The MMN to deviants with distracters in a different frequency range occurred when participants ignored the sounds by reading a book and were of similar magnitude while performing an easy one-back visual task or a more difficult three-back visual task. These results demonstrated that for auditory stimuli that can be separated on the basis of frequency, paying attention to the sounds was not necessary for the MMN to occur. This is in contrast to the case when stimuli required effort for streaming to occur (Sussman et al., 1999).

Sussman and colleagues (Sussman, Bregman, Wang, & Khan, 2005) followed up the findings by Brochard et al. (1999), which suggested a lack of streaming in multiple unattended streams, by testing whether MMN could be elicited to deviants occurring in multiple unattended streams. Tones in three separate frequency ranges (L; M, medium; and H) were presented in a repeating pattern of LMH. Within the L and M tones was a repeating pattern of triplets with ascending frequency, with occasional pattern deviants (as in Sussman et al., 1998, 1999). The H tones were presented at a constant frequency with occasional deviant frequencies, which could occur one or two times in a row. In the ignore conditions, participants watched a muted movie (Experiment 1) or performed a difficult visual task (Experiment 2) and ignored the sound stimuli. In the attend condition, participants were asked to attend to the H tones and press a button whenever they heard two deviant H tones in a row. In the ignore condition of each experiment, reliable MMN occurred following the deviants in all three streams (also see Nager, Teder-Salejarvi, Kunze, & Münte, 2003). In the attend condition, however, MMN occurred only for the attended tones, suggesting that attention to one stream actually prevented streaming of unattended tones. In another study, MMN was present (though attenuated) for two unattended tones presented at different spatial locations from the attended tones and from each other, suggesting that the unattended tones were segregated from each other (Arnott & Alain, 2002).

In a study described above, Sussman et al. (2007) compared MMN to intensity deviants early and late in the sequence and showed MMN only for deviants late in the sequence when the frequency separation between L and H tones was very large (19 semitones), providing evidence that MMN reflects buildup of stream segregation. The participants were ignoring the sound stimuli, leading the authors to further argue that buildup occurred in the absence of attention. However, this is not necessarily inconsistent with previous behavioral findings that attention affects streaming (Carlyon et al., 2001; Cusack et al., 2004) because in these behavioral studies, some buildup may have occurred for very large frequency separations even when participants ignored the tone patterns. Furthermore, it is not possible from the study by Sussman et al. (2007) to say if attention affects buildup because there was no condition when participants were paying attention with which to compare the condition when they were ignoring the stimuli.

Snyder et al. (2006) more directly examined the effects of attention on streaming by measuring sensory-evoked responses while participants were attending to ABA- patterns or ignoring them while watching a subtitled muted movie. In both attend and ignore conditions, the frequency separation-related modulations increased as a function of frequency separation (Figure 4A). In the attend condition, an additional positive enhancement of activity peaked about 200 ms after the beginning of each repetition of the ABA- pattern and increased over the course of the trial, paralleling the perceptual buildup of streaming that occurs over several seconds (Figure 4B, left column). Whereas ignoring the ABA- patterns had little effect on the frequency separation-related modulations, it markedly reduced the positive activity that increased during the course of the trial (Figure 4B, right column). This suggested that although attention had little effect on frequency-based segregation of sounds, the buildup process is highly dependent on paying attention. As discussed above, Winkler et al. (2005) also found evidence for distinct stages of processing in temporal

integration that depended on a representation of one stream, an early one that depended only on frequency separation and a later one that depended on whether listeners perceived one or two streams. How these two processes are related to those indexed by the ERPs reported by Snyder et al. (2006) requires further investigation. However, the two studies provide converging evidence for multiple stages of processing in stream segregation.

On the basis of recent evidence (Snyder et al., 2006; Winkler et al., 2005), at least two processes, frequency-based segregation and buildup, can thus be distinguished from each other and labeled according to Bregman's dichotomy as primary and schema-based processes, respectively (Bregman, 1990). This is also consistent with the hierarchical model proposed by Cusack et al. (2004), which proposed that automatic segregation of sounds occurs, but that the buildup of streaming depends on focused attention. This model further suggests that previous findings showing a lack of an attention effect on streaming may have been tapping into frequency-based segregation processes (e.g., Macken et al., 2003; Sussman et al., 1999; Winkler, Sussman, et al., 2003), whereas studies showing an attention effect may have been tapping into buildup processes (e.g., Carlyon et al., 2001; Cusack et al., 2004). An important question for future studies is whether segregation of tones differing in centrally computed attributes such as periodicity pitch also occurs automatically. This would test the generality of the conclusion that segregation does not depend on attention.

A prediction that follows from the interpretation of segregation and buildup as two distinct processes is that it would be possible to have segregation without buildup. Although this possibility has not been directly examined, it is likely that perception of integrated streams would be impossible for very slow presentation rates, even with large frequency separations. Indirect evidence in support of a lack of integration for tones presented at slow rates comes from studies showing difficulty in perceiving and producing temporal relationships between tones with very long SOAs (Engström, Kelso, & Holroyd, 1996; Friberg & Sundberg, 1995; Mates, Radil, Müller, & Pöppel, 1994; Peters, 1989).

The Role of Context

The fact that it takes several seconds for perception of streaming to build up from the beginning of a trial (Anstis & Saida, 1985; Bregman, 1978) suggests that streaming is associated with a relatively long temporal integration process. It is therefore likely that stream segregation is affected by preceding context as long as it occurs in this putative temporal integration window of several seconds. For example, presenting events prior to a repeating ABA-pattern might increase or decrease the likelihood of perceiving streaming. The idea that integration of information over several seconds may play a role in explaining effects of context on streaming is consistent with previous behavioral findings. For example, one study (Bregman, 1978) presented a pattern of four tones that alternated between two different frequency ranges. These four-tone patterns were presented repetitively with a fixed silent interval while participants manually adjusted the presentation rate between alternating tones until they could hear two segregated streams. Increasing the silent interval between successive presentations of the four-tone patterns from 0 to 4 s greatly increased the presentation rate necessary to perceive streaming,

suggesting that integration of information occurred over several seconds and increased the likelihood of streaming.

Another study tested the duration of stream biasing by using an induction sequence with tones that had the same frequency as the A tones from a repeating AB test pattern but had an SOA that was twice as fast as the A tones in the AB pattern (Beauvois & Meddis, 1997). They found that the biasing effect of the induction sequence decayed exponentially with a time constant of 7.84 s in musicians and 1.42 s in nonmusicians. Such an effect could be related to the importance of perceiving and producing musical patterns that are structured over long time spans for musicians, and it suggests a high degree of plasticity for processes related to buildup.

A recent study looked at effects of context for ABA- patterns with four different frequency separations between A and B tones intermixed in blocks of trials (Snyder et al., 2007). On each trial containing 27 ABA- repetitions, the A tone had a fixed frequency, and the B tone either had the same frequency as the A tone or one of three higher frequencies. Listeners were more likely to perceive streaming on the current trial when the previous trial had a smaller frequency separation between A and B tones, a contrastive context effect. Effects of the previous trial appeared even when listeners deliberately tried to perceive a specified organization in the current trial (i.e., one stream or two streams; cf. Van Noorden, 1975). The effect of the previous trial on the current trial diminished when increasing the intertrial intervals from 1.44 s to 5.76 s, consistent with previously reported durations of context effects (Beauvois & Meddis, 1997; Bregman, 1978). However, the effect of the previous trial (lag 1) was still present at 5.76 s. Furthermore, the trial before the previous trial (lag 2) also had an influence on the current trial. These data provide evidence that long temporal integration processes such as those observed in sensory memory (Cowan, 1984) and/or long-lasting neural suppression (Micheyl et al., 2005; Ulanovsky, Las, Farkas, & Nelken, 2004; Ulanovsky, Las, & Nelken, 2003) might underlie contrastive effects of context on streaming.

To test the stimulus generality of context effects, a study used a test sequence consisting of a repeating ABA- pattern that was preceded by one of several induction patterns: (a) an isochronous series of tones identical in frequency to the A tones presented at the same rate as the ABA- pattern but with longer duration tones; (b) an isochronous series of tones identical in frequency, duration, and rate to the A tones; (c) an irregular series of tones identical in frequency, mean duration, and mean presentation rate to the A tones; and (d) continuous white noise (Rogers & Bregman, 1993). The first three (nonnoise) induction sequences all enhanced perception of streaming in the test sequence compared with that for the induction sequence with noise. The induction sequences with overall rate similar to the A tones were slightly better at inducing streaming than was the sequence with a similar rate to the ABA-pattern. These results suggested that the induction sequences helped participants attend to the A tones separately from the B tones. Such an interpretation implies that context can activate schema that affect perception of streaming. It is also possible that frequency-specific suppression played an important role in these biasing effects (Micheyl et al., 2005). For example, suppression as indexed by the N1 response lasts for several seconds (Carver et al., 2002; Hari et al., 1982; Lu et al., 1992; Sams, Hari, Rif, & Knuutila, 1993), similar to the duration of context effects discussed above. This argument is weakened by the finding that an

induction sequence consisting of a continuous tone at the A-tone frequency did not enhance streaming in the test sequence compared with that in a white noise induction sequence (Rogers & Bregman, 1993). It is possible, however, that multiple event onsets, as opposed to a single sustained tone, are necessary for generating the suppression that underlies facilitative context effects.

A recent study used MMN to study neural correlates of the effect of prior context on streaming (Sussman & Steinschneider, 2006). The test sequence was a series of L and H tones arranged in a repeating LHHH pattern. The SOA between successive tones was 180 ms, and the frequency separation was five semitones, making the test sequence possible to hear either as one stream or two streams. Three different induction sequences immediately preceded the test sequences with no break in the rhythm: (a) only the L tones, (b) the same pattern as the test sequence but with a smaller frequency separation, and (c) the same pattern as the test sequence. Occasionally, an L tone in the test sequence had a deviant intensity. In all conditions, participants ignored the sounds by watching a movie. MMN to the deviant intensity occurred only when the induction sequence consisted of L tones, suggesting that this context enhanced segregation of the L and H tones in the test sequence. This is consistent with previous findings that patterns consisting of one of the tone frequencies in a test sequence can enhance stream segregation (Beauvois & Meddis, 1997; Bregman, 1978; Rogers & Bregman, 1993). However, the fact that the induction sequence with a smaller frequency separation than the test sequence did not also facilitate generation of MMN does not correspond with the finding that the larger the frequency separation of the previous trial, the more streaming was reported on the current trial (Snyder et al., 2007). The lack of behavioral data collected to compare with the MMN data thus makes it difficult to evaluate how closely the MMN reflects effects of context on perception of streaming.

The Role of Intention

In situations in which an individual knows what sounds they wish to attend, it is possible to intentionally bias perception to stream the target sounds from the background. On the other hand, if the individual is attempting to perceive a single pattern that varies in frequency or some other acoustic dimension, it is possible to bias perception for coherence. Van Noorden (1975) was the first to describe the effect of trying to hear one particular organization (i.e., coherence or streaming). ABA- patterns were presented in which the B tone was set at a constant frequency whereas the A tone was continuously changed throughout the trial, moving from one octave above the B tone to one octave below the B tone. On some trials, participants attempted to perceive a galloping pattern (i.e., coherence) for as much of the time as possible; on other trials, participants attempted to perceive streaming by listening selectively to the unchanging A tone. The presentation rate was varied to determine how tempo, in addition to frequency separation, affected the ability to hear a single pattern with a galloping rhythm or two parallel streams. Two different perceptual boundaries were described on the basis of the results. The *temporal coherence boundary* was defined as the frequency separation between A and B tones above which participants could no longer perceive one stream despite trying to do so. The *fission boundary* was defined

as the frequency separation below which participants could no longer perceive two streams despite trying to do so. At very fast presentation rates (e.g., 50-ms SOA), these two boundaries had similar values for each participant (two to five semitones, depending on the participant); however, at slower presentation rates (e.g., 150-ms SOA), the two thresholds diverged sharply. This was mainly due to the fact that the temporal coherence boundary increased sharply with slower presentation rates, whereas the fission boundary did not change much with presentation rate. Similar results were obtained when participants tried to hear one organization or the other while they adjusted the frequency separation themselves.

The results of Van Noorden (1975) suggested that at very fast presentation rates streaming is based largely on stimulus-driven or primary mechanisms, whereas at slower rates participants can strongly bias their judgments in one direction or another. Furthermore, it appears that a fairly constant minimum frequency separation for obligatory perception of one stream exists across a large range of presentation rates, which could provide an important clue about the frequency resolution and the processing stage at which frequency-based segregation occurs. Although this study provides important clues about both primary and schema-driven mechanisms of streaming, it also raises some important questions that have yet to be addressed empirically. Of most noted importance, it is not clear at what stage of processing intention affects stream segregation. One possibility is that intention affects frequency-based segregation by, for example, decreasing the tuning bandwidth of auditory neurons, which would increase reports of streaming. Alternatively, intention could also affect buildup. These two alternatives could be tested by using recently identified neural correlates of these processes (Micheyl et al., 2005; Snyder et al., 2006). However, it is also possible that intention could simply change response criteria, a possibility that deserves to be addressed empirically. Another important issue is which brain areas are used for intentionally perceiving one organization versus the other. It seems likely that brain areas implicated in attention such as the prefrontal and parietal cortices would play an important role, a possibility that could be tested with dipole source analysis of neuroelectric or neuromagnetic brain activity, or by using fMRI.

The Role of Knowledge

It is likely that participants use knowledge of speech, music, and environmental sounds to segregate task-relevant sounds from irrelevant ones (i.e., background). Presenting familiar material interleaved with distracting events is one way to determine whether and how people are able to use knowledge or "schema" to aid segregation. For example, Dowling (1973) presented familiar melodies (e.g., "Twinkle, Twinkle Little Star") that were interleaved with other melodies. The likelihood of identifying the familiar melody increased with larger frequency separation between the two melodies. When told the name of the target melody, identification performance was well above chance even with no frequency separation. However, knowing the name of the background melody did not help participants. This suggested that using a schema to guide segregation might act by highlighting events in the target melody rather than diminishing events in the distracter melody. In a subsequent study, this conclusion was further supported and generalized to the time domain by showing that melodies with

events presented off the beat of a global pattern formed by two interleaved melodies in the same frequency range were recognized less well than melodies with events presented on the beat (Dowling, Lung, & Herrbold, 1987). This suggests that participants expected events in the target melody to occur at specific points in frequency and time, thus aiding segregation from the distracter melody (cf. M. R. Jones, Kidd, & Wetzell, 1981; M. R. Jones, Moynihan, MacKenzie, & Puente, 2002).

Instead of using familiar melodies stored in long-term memory, another study used unfamiliar melodies that were stored in short-term memory to test for effects of schema on segregation (Bey & McAdams, 2002). This study presented two unfamiliar melodies to participants on each trial, one after the other. Sometimes the two melodies were the same and sometimes they were slightly different, with the task being to report whether there was a change from one melody to the other. One of the melodies was interleaved with distracting tones that, to varying degrees, could overlap in pitch range with the target melody. On half of the trials, the interleaved melody was presented first; on the other half of the trials, the noninterleaved melody was presented first. When the noninterleaved melody was presented first, thus providing participants with a template for picking out the target melody, performance was markedly improved. However, such an improvement did not occur when the two melodies had the same frequency range. This is in contrast to Dowling (1973), who showed improved performance for identifying familiar melodies interleaved with distracter melodies in the same frequency range. The discrepancy between these studies could be due to participants relying on short-term memory versus long-term memory. However, another possibility is that the discrepancy arises not from a difference in memory systems per se but rather from a difference in the strength of the representation. This rests on the assumption that highly familiar melodies are likely to have much stronger representations than do melodies that were just heard for the first time. Future studies could therefore compare brain activity during schema-based segregation with familiar versus just-learned melodies, or sentences, to explain the different performance in these two paradigms.

An important question that could be addressed by using measurements of brain activity during schema-based segregation is the extent to which feedback signals are important in streaming. For example, one might hypothesize that schematic information (e.g., familiar music or native speech sounds) stored in higher level brain areas feeds back to brain areas that perform basic aspects of segregation. On the other hand, it is also possible that brain areas that store schematic information are able to carry out segregation on complex signals such as music and speech with only feed-forward information from lower levels. Thus, schema-based segregation of sounds may provide an important model for understanding whether and how knowledge interacts with low-level perceptual processing.

Bistability of Streaming and "Voting"

An overlooked aspect of stream segregation is the possibility that it represents a case of auditory bistability, similar to many visual patterns that can be perceived in one of two qualitatively different organizations (for a review of visual bistability, see Leopold & Logothetis, 1999). The bistable nature of streaming is often not considered because experiments often use short se-

quences of tones resulting in perception of one stream during the buildup period followed by a single switch to perceiving two streams. Pressnitzer and Hupé (2006) studied perception of streaming for long ABA- sequences and compared their results directly to perception of a visual moving plaid stimulus that has bistable properties. The results showed that both the auditory and visual patterns were initially perceived as one object (i.e., a single galloping stream or a single moving plaid stimulus, respectively) during the buildup phase before switching to two objects (i.e., a high and a low pattern of tones or two gratings moving past each other, respectively), followed by alternating perception between one and two objects.

The fact that streaming can be classified as a bistable perceptual phenomenon implies that an important aspect of streaming, in addition to the segregation and buildup processes, is how the nervous system decides at any point in time what is the perceptual experience of the listener (i.e., "voting"). The necessity of including voting in theories of streaming suggests that previous models that do not include voting are incomplete. For example, the model by Micheyl et al. (2005) that relies on unidirectional suppression toward a fixed threshold for deciding whether a neuron is coding for one or two sounds may account only for the initial perception of one stream and the first switch to two streams, without being able to account for subsequent switches back and forth between perception of one and two streams. Thus, it may be appropriate to generalize the notion of buildup to also include the time it takes for perception to switch back to one stream after the initial switch from one to two streams, in addition to subsequent switches between one and two streams. However, it would be important to test whether buildup from one to two streams has similar behavioral characteristics and neural correlates as buildup from two streams to one. Another implication of including voting as an aspect of streaming is that it offers another stage of processing at which attention could affect streaming, apart from any effects of attention on segregation or buildup.

On the basis of the bistable nature of stream segregation and on results suggesting multiple levels of processing in streaming, Denham and Winkler (2006) proposed a model of streaming. Their model consisted of (a) segregation that incorporated synaptic depression as a mechanism for the buildup of streaming, (b) predictive modeling that implements integration of events within a stream, (c) competition between alternative organizations (e.g., one stream vs. two streams), and (d) adaptation of the current perceptual organization to account for the switches that occur between organizations. It is important to evaluate and build on such models to develop further hypotheses about the neurophysiological mechanisms of various aspects of streaming.

General Conclusions

The studies reviewed in this article used a number of different methodologies, including purely psychophysical methods, neurophysiological measurements in humans and nonhuman animals, and computational models that incorporated basic aspects of auditory anatomy and physiology. Together, these studies suggest an emerging picture of when, where, how, and what the auditory system organizes during auditory stream segregation. On the surface, auditory stream segregation is a relatively simple phenomenon, with this simplicity being a prime reason for its usefulness in

understanding auditory perceptual organization. The studies reviewed here, however, suggest a high degree of complexity in the range of mechanisms and the number of processing levels that appear to be involved in different aspects of streaming.

Here, we propose a general theoretical framework for understanding the findings we reviewed above, which we hope will serve as a guide for future investigations:

1. *Streaming is a process that occurs at multiple levels of the auditory system.* This idea stems from the facts that (a) the auditory system contains many different stages of processing; (b) stream segregation can occur on the basis of many different acoustic cues, some of which are extracted at the earliest stages of auditory processing and some of which are extracted only after central computations, such as those required for integration across the two ears or across multiple frequency regions; and (c) stream segregation involves not only the precise coding of stimulus features but also more general perceptual and cognitive processes, such as attention, integration of prior context, intention, schematic knowledge, and perceptual competition. Thus, a major goal for future studies should be to identify neural processes and substrates for the extraction of different acoustic cues to streaming in addition to these higher order processes. Understanding how different processing stages interact with each other through feed-forward and feedback connections is a related goal of great importance.

2. *Neural suppression plays an important role in segregation and buildup.* The evidence from invasive neurophysiological studies in animals suggests that suppression on both short-term and longer term scales is important for segregation and buildup. Therefore, an important goal is to identify specific mechanisms of suppression that are important for streaming. The clearest candidate for short-term suppression is GABAergic inhibition, especially for sequentially presented tones separated by less than 200 ms. For longer term suppression, such as the gradual decrease in response amplitude for long tone sequences, synaptic depression is a likely candidate.

3. *Attention influences some processes more so than others.* Although there is now evidence for differential effects of attention on early versus late stages of stream segregation, more evidence is necessary to confirm this and to identify precisely the extent to which different aspects of streaming are affected by attention.

4. *Stream segregation involves competition between alternative percepts.* Recent evidence on the bistable nature of streaming suggests that whether a participant hears one stream or two streams is not simply a matter of the stimulus characteristics and the amount of time that has past since the beginning of the sequence. Rather, streaming may be a dynamic process by which representations for different perceptual solutions compete. Better understanding of how the auditory system chooses whether perception will consist of one stream or two streams will likely inform mechanisms of perception with implications for other perceptual modalities, such as vision. For example, it is likely that similar mechanisms govern perceptual competition in different modalities.

5. *Are there cross-modal influences on streaming?* Given the evidence for similarities between streaming and visual bistable perception, it is possible that perception in another modality could influence perception of streaming. It is also possible that stimulus events in other modalities could influence the low-level processing of auditory attributes during stream segregation. For example, presenting visual-spatial information may enhance auditory

streaming based on auditory spatial cues. This is a topic of study that has not been addressed but could reveal important information about how perceptual segregation occurs in naturalistic situations such as a noisy urban setting or a cocktail party, in which information from multiple modalities might be used to segregate and group acoustic events into auditory objects and streams.

Although it is not yet possible to propose a more detailed and definitive account of how the nervous system accomplishes streaming, such an account will likely include multiple neurocomputational principles (e.g., tonotopy, suppression, temporal integration, competition between representations) and multiple levels of processing in the peripheral and central auditory systems, in addition to modality nonspecific regions involved in attention and perceptual organization. This mirrors an emerging view of the auditory system more generally as containing multiple processing pathways with both feed-forward and feedback connections (Kaas & Hackett, 2000), with such complex functions as attention-dependent modulation, cross-modal integration, and experience-dependent plasticity occurring as early as primary and secondary auditory cortex (DeWeese, Hromadka, & Zador, 2005). To illuminate how streaming fits into these new perspectives on the auditory system, future studies should investigate how sounds are coded at multiple levels of the auditory system, including subcortical areas. Similarly, understanding how the range of stimulus features that give rise to streaming are coded at the various stages of auditory processing and how cognitive variables such as attention, context, and intention influence processing at these stages are important goals for future studies. Finally, in both human and animal studies, it is vital to obtain behavioral responses in addition to neurophysiological data to determine the extent to which the observed neural correlates of streaming are related to stimulus coding versus perception of streaming per se. Thus, whereas a comprehensive mechanistic theory of auditory stream segregation must wait for more data, it is clear in many cases what studies should be done next to attain a more detailed understanding of streaming.

References

- Akeroyd, M. A., Carlyon, R. P., & Deeks, J. M. (2005). Can dichotic pitches form two streams? *Journal of the Acoustical Society of America*, *118*, 977–981.
- Alain, C. (in press). Breaking the wave: Effects of attention and learning on concurrent sound perception. *Hearing Research*.
- Alain, C., Achim, A., & Richer, F. (1993). Perceptual context and the selective attention effect on auditory event-related brain potentials. *Psychophysiology*, *30*, 572–580.
- Alain, C., Achim, A., & Woods, D. L. (1999). Separate memory-related processing for auditory frequency and patterns. *Psychophysiology*, *36*, 737–744.
- Alain, C., & Arnott, S. R. (2000). Selectively attending to auditory objects. *Frontiers in Bioscience*, *5*, D202–D212.
- Alain, C., & Izenberg, A. (2003). Effects of attentional load on auditory scene analysis. *Journal of Cognitive Neuroscience*, *15*, 1063–1073.
- Alain, C., Ogawa, K. H., & Woods, D. L. (1996). Aging and the segregation of auditory stimulus sequences. *Journals of Gerontology, Series B: Psychological Sciences and Social Sciences*, *51*, P91–P93.
- Alain, C., & Woods, D. L. (1993). Distractor clustering enhances detection speed and accuracy during selective listening. *Perception & Psychophysics*, *54*, 509–514.

- Alain, C., & Woods, D. L. (1994). Signal clustering modulates auditory cortical activity in humans. *Perception & Psychophysics*, *56*, 501–516.
- Anstis, S., & Saida, S. (1985). Adaptation to auditory streaming of frequency-modulated tones. *Journal of Experimental Psychology: Human Perception and Performance*, *11*, 257–271.
- Arnott, S. R., & Alain, C. (2002). Effects of perceptual context on event-related brain potentials during auditory spatial attention. *Psychophysiology*, *39*, 625–632.
- Barker, B. A., & Newman, R. S. (2004). Listen to your mother! The role of talker familiarity in infant streaming. *Cognition*, *94*, B45–B53.
- Beauvois, M. W., & Meddis, R. (1996). Computer simulation of auditory stream segregation in alternating-tone sequences. *Journal of the Acoustical Society of America*, *99*, 2270–2280.
- Beauvois, M. W., & Meddis, R. (1997). Time decay of auditory stream biasing. *Perception & Psychophysics*, *59*, 81–86.
- Bee, M. A., & Klump, G. M. (2004). Primitive auditory stream segregation: A neurophysiological study in the songbird forebrain. *Journal of Neurophysiology*, *92*, 1088–1104.
- Bey, C., & McAdams, S. (2002). Schema-based processing in auditory scene analysis. *Perception & Psychophysics*, *64*, 844–854.
- Bregman, A. S. (1978). Auditory streaming is cumulative. *Journal of Experimental Psychology: Human Perception and Performance*, *4*, 380–387.
- Bregman, A. S. (1990). *Auditory scene analysis: The perceptual organization of sound*. Cambridge, MA: MIT Press.
- Bregman, A. S., Ahad, P. A., Crum, P. A. C., & O'Reilly, J. (2000). Effects of time intervals and tone durations on auditory stream segregation. *Perception & Psychophysics*, *62*, 626–636.
- Bregman, A. S., & Campbell, J. (1971). Primary auditory stream segregation and perception of order in rapid sequences of tones. *Journal of Experimental Psychology*, *89*, 244–249.
- Brochard, R., Drake, C., Botte, M. C., & McAdams, S. (1999). Perceptual organization of complex auditory sequences: Effect of number of simultaneous subsequences and frequency separation. *Journal of Experimental Psychology: Human Perception and Performance*, *25*, 1742–1759.
- Butler, R. A. (1968). Effect of changes in stimulus frequency and intensity on habituation of human vertex potential. *Journal of the Acoustical Society of America*, *44*, 945–950.
- Carlyon, R. P. (2004). How the brain separates sounds. *Trends in Cognitive Sciences*, *8*, 465–471.
- Carlyon, R. P., Cusack, R., Foxton, J. M., & Robertson, I. H. (2001). Effects of attention and unilateral neglect on auditory stream segregation. *Journal of Experimental Psychology: Human Perception and Performance*, *27*, 115–127.
- Carlyon, R. P., Plack, C. J., Fantini, D. A., & Cusack, R. (2003). Cross-modal and non-sensory influences on auditory streaming. *Perception*, *32*, 1393–1402.
- Carver, F. W., Fuchs, A., Jantzen, K. J., & Kelso, J. A. S. (2002). Spatiotemporal analysis of the neuromagnetic response to rhythmic auditory stimulation: Rate dependence and transient to steady-state transition. *Clinical Neurophysiology*, *113*, 1921–1931.
- Chatterjee, M., Sarampalis, A., & Oba, S. I. (2006). Auditory stream segregation with cochlear implants: A preliminary report. *Hearing Research*, *222*, 100–107.
- Chen, Q. C., & Jen, P. H. (2000). Bicuculline application affects discharge patterns, rate-intensity functions, and frequency tuning characteristics of bat auditory cortical neurons. *Hearing Research*, *150*, 161–174.
- Cherry, E. C. (1953). Some experiments on the recognition of speech, with one and with two ears. *Journal of the Acoustical Society of America*, *25*, 975–979.
- Cooper, H. R., & Roberts, B. (2007). Auditory stream segregation of tone sequences in cochlear implant listeners. *Hearing Research*, *225*, 11–24.
- Cowan, N. (1984). On short and long auditory stores. *Psychological Bulletin*, *96*, 341–370.
- Cusack, R. (2005). The intraparietal sulcus and perceptual organization. *Journal of Cognitive Neuroscience*, *17*, 641–651.
- Cusack, R., Deeks, J., Aikman, G., & Carlyon, R. P. (2004). Effects of location, frequency region, and time course of selective attention on auditory scene analysis. *Journal of Experimental Psychology: Human Perception and Performance*, *30*, 643–656.
- Cusack, R., & Roberts, B. (1999). Effects of similarity in bandwidth on the auditory sequential streaming of two-tone complexes. *Perception*, *28*, 1281–1289.
- Cusack, R., & Roberts, B. (2000). Effects of differences in timbre on sequential grouping. *Perception & Psychophysics*, *62*, 1112–1120.
- Darwin, C. J. (1997). Auditory grouping. *Trends in Cognitive Sciences*, *1*, 327–333.
- Davis, H., Mast, T., Yoshie, N., & Zerlin, S. (1966). The slow response of the human cortex to auditory stimuli: Recovery process. *Electroencephalography and Clinical Neurophysiology*, *21*, 105–113.
- Deike, S., Gaschler-Markefski, B., Brechmann, A., & Scheich, H. (2004). Auditory stream segregation relying on timbre involves left auditory cortex. *NeuroReport*, *15*, 1511–1514.
- Demany, L. (1982). Auditory stream segregation in infancy. *Infant Behavior & Development*, *5*, 261–276.
- Denham, S. L., & Winkler, I. (2006). The role of predictive models in the formation of auditory streams. *Journal of Physiology (Paris)*, *100*, 154–170.
- DeWeese, M. R., Hromadka, T., & Zador, A. M. (2005). Reliability and representational bandwidth in the auditory cortex. *Neuron*, *48*, 479–488.
- Dowling, W. J. (1973). Perception of interleaved melodies. *Cognitive Psychology*, *5*, 322–337.
- Dowling, W. J., Lung, K. M. T., & Herrbold, S. (1987). Aiming attention in pitch and time in the perception of interleaved melodies. *Perception & Psychophysics*, *41*, 642–656.
- Dutar, P., & Nicoll, R. A. (1988, March 10). A physiological role for GABA_B receptors in the central nervous system. *Nature*, *332*, 156–158.
- Engström, D. A., Kelso, J. A. S., & Holroyd, T. (1996). Reaction-anticipation transitions in human perception-action patterns. *Human Movement Science*, *15*, 809–822.
- Fay, R. R. (1998). Auditory stream segregation in goldfish (*Carassius auratus*). *Hearing Research*, *120*, 69–76.
- Fay, R. R. (2000). Spectral contrasts underlying auditory stream segregation in goldfish (*Carassius auratus*). *Journal of the Association for Research in Otolaryngology*, *1*, 120–128.
- Feng, A. S., & Ratnam, R. (2000). Neural basis of hearing in real-world situations. *Annual Review of Psychology*, *51*, 699–725.
- Fishman, Y. I., Arezzo, J. C., & Steinschneider, M. (2004). Auditory stream segregation in monkey auditory cortex: Effects of frequency separation, presentation rate, and tone duration. *Journal of the Acoustical Society of America*, *116*, 1656–1670.
- Fishman, Y. I., Reser, D. H., Arezzo, J. C., & Steinschneider, M. (2001). Neural correlates of auditory stream segregation in primary auditory cortex of the awake monkey. *Hearing Research*, *151*, 167–187.
- Formisano, E., Kim, D. S., Di Salle, F., van de Moortele, P. F., Ugurbil, K., & Goebel, R. (2003). Mirror-symmetric tonotopic maps in human primary auditory cortex. *Neuron*, *40*, 859–869.
- Friberg, A., & Sundberg, J. (1995). Time discrimination in a monotonic, isochronous sequence. *Journal of the Acoustical Society of America*, *98*, 2524–2531.
- Godey, B., Schwartz, D., de Graaf, J. B., Chauvel, P., & Liegeois-Chauvel, C. (2001). Neuromagnetic source localization of auditory evoked fields and intracerebral evoked potentials: A comparison of data in the same patients. *Clinical Neurophysiology*, *112*, 1850–1859.
- Grimault, N., Bacon, S. P., & Micheyl, C. (2002). Auditory stream segregation on the basis of amplitude-modulation rate. *Journal of the Acoustical Society of America*, *111*, 1340–1348.
- Grimault, N., Micheyl, C., Carlyon, R. P., Arthaud, P., & Collet, L. (2000).

- Influence of peripheral resolvability on the perceptual segregation of harmonic complex tones differing in fundamental frequency. *Journal of the Acoustical Society of America*, 108, 263–271.
- Grimault, N., Micheyl, C., Carlyon, R. P., Arthaud, P., & Collet, L. (2001). Perceptual auditory stream segregation of sequences of complex sounds in subjects with normal and impaired hearing. *British Journal of Audiology*, 35, 173–182.
- Gutschalk, A., Melcher, J. R., Micheyl, C., Wilson, C., & Oxenham, A. (2006). Neural correlates of streaming without spectral cues in human auditory cortex. *Association for Research in Otolaryngology Abstracts*, 172.
- Gutschalk, A., Micheyl, C., Melcher, J. R., Rupp, A., Scherg, M., & Oxenham, A. J. (2005). Neuromagnetic correlates of streaming in human auditory cortex. *Journal of Neuroscience*, 25, 5382–5388.
- Hansen, J. C., & Hillyard, S. A. (1980). Endogenous brain potentials associated with selective auditory attention. *Electroencephalography and Clinical Neurophysiology*, 49, 277–290.
- Hansen, J. C., & Hillyard, S. A. (1988). Temporal dynamics of human auditory selective attention. *Psychophysiology*, 25, 316–329.
- Hari, R., Kaila, K., Katila, T., Tuomisto, T., & Varpula, T. (1982). Interstimulus-interval dependence of the auditory vertex response and its magnetic counterpart: Implications for their neural generation. *Electroencephalography and Clinical Neurophysiology*, 54, 561–569.
- Hartmann, W. M., & Johnson, D. (1991). Stream segregation and peripheral channeling. *Music Perception*, 9, 155–184.
- Hillyard, S. A., Hink, R. F., Schwent, V. L., & Picton, T. W. (1973, October 12). Electrical signs of selective attention in the human brain. *Science*, 182, 171–180.
- Hollich, G., Newman, R. S., & Jusczyk, P. W. (2005). Infants' use of synchronized visual information to separate streams of speech. *Child Development*, 76, 598–613.
- Hong, R. S., & Turner, C. W. (2006). Pure-tone auditory stream segregation and speech perception in noise in cochlear implant recipients. *Journal of the Acoustical Society of America*, 120, 360–374.
- Howard, M. A., Volkov, I. O., Abbas, P. J., Damasio, H., Ollendieck, M. C., & Granner, M. A. (1996). A chronic microelectrode investigation of the tonotopic organization of human auditory cortex. *Brain Research*, 724, 260–264.
- Hulse, S. H. (2002). Auditory scene analysis in animal communication. In P. J. B. Slater, J. S. Rosenblatt, C. T. Snowdon, & T. J. Roper (Eds.), *Advances in the study of behavior* (Vol. 31, pp. 163–200). San Diego, CA: Academic Press.
- Hulse, S. H., MacDougall-Shackleton, S. A., & Wisniewski, A. B. (1997). Auditory scene analysis by songbirds: Stream segregation of birdsong by European starlings (*Sturnus vulgaris*). *Journal of Comparative Psychology*, 111, 3–13.
- Hung, J., Jones, S. J., & Vaz Pato, M. (2001). Scalp potentials to pitch change in rapid tone sequences: A correlate of sequential stream segregation. *Experimental Brain Research*, 140, 56–65.
- Izumi, A. (2002). Auditory stream segregation in Japanese monkeys. *Cognition*, 82, B113–B122.
- Jones, M. R., Kidd, G., & Wetzell, R. (1981). Evidence for rhythmic attention. *Journal of Experimental Psychology: Human Perception and Performance*, 7, 1059–1073.
- Jones, M. R., Moynihan, H., MacKenzie, N., & Puente, J. (2002). Temporal aspects of stimulus-driven attending in dynamic arrays. *Psychological Science*, 13, 313–319.
- Jones, S. J., Longe, O., & Vaz Pato, M. (1998). Auditory evoked potentials to abrupt pitch and timbre change of complex tones: Electrophysiological evidence of 'streaming'? *Electroencephalography and Clinical Neurophysiology*, 108, 131–142.
- Kaas, J. H., & Hackett, T. A. (2000). Subdivisions of auditory cortex and processing streams in primates. *Proceedings of the National Academy of Sciences of the United States of America*, 97, 11793–11799.
- Kanwal, J. S., Medvedev, A. V., & Micheyl, C. (2003). Neurodynamics for auditory stream segregation: Tracking sounds in the mustached bat's natural environment. *Network: Computation in Neural Systems*, 14, 413–435.
- Langner, G. (1997). Neural processing and representation of periodicity pitch. *Acta Oto Laryngologica – Supplement*, 532, 68–76.
- Leopold, D. A., & Logothetis, N. K. (1999). Multistable phenomena: Changing views in perception. *Trends in Cognitive Sciences*, 3, 254–264.
- Logothetis, N. K., & Wandell, B. A. (2004). Interpreting the BOLD signal. *Annual Review of Physiology*, 66, 735–769.
- Lu, Z. L., Williamson, S. J., & Kaufman, L. (1992, December 4). Behavioral lifetime of human auditory sensory memory predicted by physiological measures. *Science*, 258, 1668–1670.
- MacDougall-Shackleton, S. A., Hulse, S. H., Gentner, T. Q., & White, W. (1998). Auditory scene analysis by European starlings (*Sturnus vulgaris*): Perceptual segregation of tone sequences. *Journal of the Acoustical Society of America*, 103, 3581–3587.
- Macken, W. J., Tremblay, S., Houghton, R. J., Nicholls, A. P., & Jones, D. M. (2003). Does auditory streaming require attention? Evidence from attentional selectivity in short-term memory. *Journal of Experimental Psychology: Human Perception and Performance*, 29, 43–51.
- Martin, B. A., & Boothroyd, A. (2000). Cortical, auditory, evoked potentials in response to changes of spectrum and amplitude. *Journal of the Acoustical Society of America*, 107, 2155–2161.
- Mates, J., Radil, T., Müller, U., & Pöppel, E. (1994). Temporal integration in sensorimotor synchronization. *Journal of Cognitive Neuroscience*, 6, 332–340.
- McAdams, S., & Bertoncini, J. (1997). Organization and discrimination of repeating sound sequences by newborn infants. *Journal of the Acoustical Society of America*, 102, 2945–2953.
- McAlpine, D. (2004). Neural sensitivity to periodicity in the inferior colliculus: Evidence for the role of cochlear distortions. *Journal of Neurophysiology*, 92, 1295–1311.
- McCabe, S. L., & Denham, M. J. (1997). A model of auditory streaming. *Journal of the Acoustical Society of America*, 101, 1611–1621.
- McCormick, D. A. (1989). GABA as an inhibitory neurotransmitter in human cerebral cortex. *Journal of Neurophysiology*, 62, 1018–1027.
- McKenzie, D. N., & Barry, R. J. (2006). The independence of memory traces of attended and unattended stimuli. *Cerebral Cortex*, 16, 1566–1570.
- Merzenich, M. M., & Brugge, J. F. (1973). Representation of the cochlear partition of the superior temporal plane of the macaque monkey. *Brain Research*, 50, 275–296.
- Micheyl, C., Carlyon, R. P., Gutschalk, A., Melcher, J. R., Oxenham, A. J., Rauschecker, J. P., et al. (in press). The role of auditory cortex in the formation of auditory streams. *Hearing Research*.
- Micheyl, C., Tian, B., Carlyon, R. P., & Rauschecker, J. P. (2005). Perceptual organization of tone sequences in the auditory cortex of awake macaques. *Neuron*, 48, 139–148.
- Miller, G. A., & Heise, G. A. (1950). The trill threshold. *Journal of the Acoustical Society of America*, 22, 637–638.
- Møller, A. R. (2006). *Hearing: Anatomy, physiology, and disorders of the auditory system* (2nd ed.). Burlington, MA: Academic Press.
- Moore, B. C. J., & Gockel, H. (2002). Factors influencing sequential stream segregation. *Acta Acustica united with Acustica*, 88, 320–333.
- Moss, C. F., & Surlykke, A. (2001). Auditory scene analysis by echolocation in bats. *Journal of the Acoustical Society of America*, 110, 2207–2226.
- Munakata, Y., Casey, B. J., & Diamond, A. (2004). Developmental cognitive neuroscience: Progress and potential. *Trends in Cognitive Sciences*, 8, 122–128.
- Näätänen, R., Paavilainen, P., Tiitinen, H., Jiang, D., & Alho, K. (1993). Attention and mismatch negativity. *Psychophysiology*, 30, 436–450.
- Näätänen, R., & 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.
- Näätänen, R., Sams, M., Alho, K., Paavilainen, P., Reinikainen, K., &

- Sokolov, E. N. (1988). Frequency and location specificity of the human vertex N1 wave. *Electroencephalography and Clinical Neurophysiology*, *69*, 523–531.
- Näätänen, R., & Winkler, I. (1999). The concept of auditory stimulus representation in cognitive neuroscience. *Psychological Bulletin*, *125*, 826–859.
- Nager, W., Teder-Salejarvi, W., Kunze, S., & Münte, T. F. (2003). Preattentive evaluation of multiple perceptual streams in human audition. *NeuroReport*, *14*, 871–874.
- Newman, R. S. (2005). The cocktail party effect in infants revisited: Listening to one's name in noise. *Developmental Psychology*, *41*, 352–362.
- Newman, R. S., & Juszczyk, P. W. (1996). The cocktail party effect in infants. *Perception & Psychophysics*, *58*, 1145–1156.
- Peters, M. (1989). The relationship between variability of intertap intervals and interval duration. *Psychological Research*, *51*, 38–42.
- Picton, T. W., Alain, C., Otten, L., Ritter, W., & Achim, A. (2000). Mismatch negativity: Different water in the same river. *Audiology and Neuro-Otology*, *5*, 111–139.
- Picton, T. W., Alain, C., Woods, D. L., John, M. S., Scherg, M., Valdes-Sosa, P., et al. (1999). Intracerebral sources of human auditory-evoked potentials. *Audiology and Neuro-Otology*, *4*, 64–79.
- Picton, T. W., Woods, D. L., & Proulx, G. B. (1978). Human auditory sustained potentials. II. Stimulus relationships. *Electroencephalography and Clinical Neurophysiology*, *45*, 198–210.
- Posner, M. I., & DiGirolamo, G. J. (2000). Cognitive neuroscience: Origins and promise. *Psychological Bulletin*, *126*, 873–889.
- Pressnitzer, D., & Hupé, J. M. (2006). Temporal dynamics of auditory and visual bistability reveal common principles of perceptual organization. *Current Biology*, *16*, 1351–1357.
- Pressnitzer, D., Micheyl, C., Sayles, M., & Winter, I. M. (2007). Responses to long-duration tone sequences in the cochlear nucleus. *Association for Research in Otolaryngology Abstracts*, 131.
- Rademacher, J., Morosan, P., Schormann, T., Schleicher, A., Werner, C., Freund, H. J., et al. (2001). Probabilistic mapping and volume measurement of human primary auditory cortex. *NeuroImage*, *13*, 669–683.
- Rauschecker, J. P. (1998). Cortical processing of complex sounds. *Current Opinion in Neurobiology*, *8*, 516–521.
- Ritter, W., Sussman, E., & Molholm, S. (2000). Evidence that the mismatch negativity system works on the basis of objects. *NeuroReport*, *11*, 61–63.
- Rivier, F., & Clarke, S. (1997). Cytochrome oxidase, acetylcholinesterase, and NADPH-diaphorase staining in human supratemporal and insular cortex: Evidence for multiple auditory areas. *NeuroImage*, *6*, 288–304.
- Roberts, B., Glasberg, B. R., & Moore, B. C. J. (2002). Primitive stream segregation of tone sequences without differences in fundamental frequency or passband. *Journal of the Acoustical Society of America*, *112*, 2074–2085.
- Rogers, W. L., & Bregman, A. S. (1993). An experimental evaluation of three theories of auditory stream segregation. *Perception & Psychophysics*, *53*, 179–189.
- Rogers, W. L., & Bregman, A. S. (1998). Cumulation of the tendency to segregate auditory streams: Resetting by changes in location and loudness. *Perception & Psychophysics*, *60*, 1216–1227.
- Rose, J. E., Hind, J. E., Anderson, D. J., & Brugge, J. F. (1971). Some effects of stimulus intensity on response of auditory nerve fibers in the squirrel monkey. *Journal of Neurophysiology*, *34*, 685–699.
- Rose, M. M., & Moore, B. C. J. (1997). Perceptual grouping of tone sequences by normally hearing and hearing-impaired listeners. *Journal of the Acoustical Society of America*, *102*, 1768–1778.
- Rose, M. M., & Moore, B. C. J. (2000). Effects of frequency and level on auditory stream segregation. *Journal of the Acoustical Society of America*, *108*, 1209–1214.
- Sams, M., Hari, R., Rif, J., & Knuutila, J. (1993). The human auditory sensory memory trace persists about 10 sec: Neuromagnetic evidence. *Journal of Cognitive Neuroscience*, *5*, 363–370.
- Scherg, M., Vajsar, J., & Picton, T. W. (1989). A source analysis of the late human auditory evoked potentials. *Journal of Cognitive Neuroscience*, *1*, 336–355.
- Schul, J., & Sheridan, R. A. (2006). Auditory stream segregation in an insect. *Neuroscience*, *138*, 1–4.
- Schulze, H., & Langner, G. (1997). Periodicity coding in the primary auditory cortex of the mongolian gerbil (*Meriones unguiculatus*): Two different coding strategies for pitch and rhythm? *Journal of Comparative Physiology A*, *181*, 651–663.
- Semple, M. N., & Scott, B. H. (2003). Cortical mechanisms in hearing. *Current Opinion in Neurobiology*, *13*, 167–173.
- Shinozaki, N., Yabe, H., Sato, Y., Sutoh, T., Hiruma, T., Nashida, T., et al. (2000). Mismatch negativity (MMN) reveals sound grouping in the human brain. *NeuroReport*, *11*, 1597–1601.
- Singh, P. G., & Bregman, A. S. (1997). The influence of different timbre attributes on the perceptual segregation of complex-tone sequences. *Journal of the Acoustical Society of America*, *102*, 1943–1952.
- Snyder, J. S., & Alain, C. (2007). Sequential auditory scene analysis is preserved in normal aging adults. *Cerebral Cortex*, *17*, 501–512.
- Snyder, J. S., Alain, C., & Picton, T. W. (2006). Effects of attention on neuroelectric correlates of auditory stream segregation. *Journal of Cognitive Neuroscience*, *18*, 1–13.
- Snyder, J. S., Carter, O. L., Lee, S.-K., Hannon, E. E., & Alain, C. (2007). Effects of context on auditory stream segregation. *Association for Research in Otolaryngology Abstracts*, 314.
- Snyder, J. S., & Large, E. W. (2004). Tempo dependence of middle- and long-latency auditory responses: Power and phase modulation of the EEG at multiple time-scales. *Clinical Neurophysiology*, *115*, 1885–1895.
- Stainsby, T. H., Moore, B. C. J., Medland, P. J., & Glasberg, B. R. (2004). Sequential streaming and effective level differences due to phase-spectrum manipulations. *Journal of the Acoustical Society of America*, *115*, 1665–1673.
- Sussman, E. S. (2005). Integration and segregation in auditory scene analysis. *Journal of the Acoustical Society of America*, *117*, 1285–1298.
- Sussman, E. S., Bregman, A. S., Wang, W. J., & Khan, F. J. (2005). Attentional modulation of electrophysiological activity in auditory cortex for unattended sounds within multistream auditory environments. *Cognitive Affective and Behavioral Neuroscience*, *5*, 93–110.
- Sussman, E., Horváth, J., Winkler, I., & Orr, M. (2007). The role of attention in the formation of auditory streams. *Perception & Psychophysics*, *69*, 136–152.
- Sussman, E., Ritter, W., & Vaughan, H. G., Jr. (1998). Attention affects the organization of auditory input associated with the mismatch negativity system. *Brain Research*, *789*, 130–138.
- Sussman, E., Ritter, W., & Vaughan, H. G., Jr. (1999). An investigation of the auditory streaming effect using event-related brain potentials. *Psychophysiology*, *36*, 22–34.
- Sussman, E., & Steinschneider, M. (2006). Neurophysiological evidence for context-dependent encoding of sensory input in human auditory cortex. *Brain Research*, *1075*, 165–174.
- Sweet, R. A., Dorph-Petersen, K. A., & Lewis, D. A. (2005). Mapping auditory core, lateral belt, and parabelt cortices in the human superior temporal gyrus. *Journal of Comparative Neurology*, *491*, 270–289.
- Tan, A. Y., Zhang, L. I., Merzenich, M. M., & Schreiner, C. E. (2004). Tone-evoked excitatory and inhibitory synaptic conductances of primary auditory cortex neurons. *Journal of Neurophysiology*, *92*, 630–643.
- Tervaniemi, M., Saarinen, J., Paavilainen, P., Danilova, N., & Näätänen, R. (1994). Temporal integration of auditory information in sensory memory as reflected by the mismatch negativity. *Biological Psychology*, *38*, 157–167.
- Trainor, L. J., McDonald, K. L., & Alain, C. (2002). Automatic and controlled processing of melodic contour and interval information measured by electrical brain activity. *Journal of Cognitive Neuroscience*, *14*, 430–442.
- Trainor, L. J., & Trehub, S. E. (1989). Aging and auditory temporal sequencing: Ordering the elements of repeating tone patterns. *Perception & Psychophysics*, *45*, 417–426.

- Ulanovsky, N., Las, L., Farkas, D., & Nelken, I. (2004). Multiple time scales of adaptation in auditory cortex neurons. *Journal of Neuroscience*, *24*, 10440–10453.
- Ulanovsky, N., Las, L., & Nelken, I. (2003). Processing of low-probability sounds by cortical neurons. *Nature Neuroscience*, *6*, 391–398.
- Van Noorden, L. (1975). *Temporal coherence in the perception of tone sequences*. Unpublished doctoral dissertation, Eindhoven University of Technology, Eindhoven, the Netherlands.
- Vliegen, J., Moore, B. C. J., & Oxenham, A. J. (1999). The role of spectral and periodicity cues in auditory stream segregation, measured using a temporal discrimination task. *Journal of the Acoustical Society of America*, *106*, 938–945.
- Vliegen, J., & Oxenham, A. J. (1999). Sequential stream segregation in the absence of spectral cues. *Journal of the Acoustical Society of America*, *105*, 339–346.
- Wang, J., McFadden, S. L., Caspary, D., & Salvi, R. (2002). Gamma-aminobutyric acid circuits shape response properties of auditory cortex neurons. *Brain Research*, *944*, 219–231.
- Warren, R. M., Obusek, C. J., Farmer, R. M., & Warren, R. P. (1969, May 2). Auditory sequence: Confusion of patterns other than speech or music. *Science*, *164*, 586–587.
- Wehr, M., & Zador, A. M. (2003, November 27). Balanced inhibition underlies tuning and sharpens spike timing in auditory cortex. *Nature*, *426*, 442–446.
- Wehr, M., & Zador, A. M. (2005). Synaptic mechanisms of forward suppression in rat auditory cortex. *Neuron*, *47*, 437–445.
- Wilson, E. C., Melcher, J. R., Micheyl, C., Gutschalk, A., & Oxenham, A. J. (2007). Cortical fMRI activation to sequences of tones alternating in frequency: Relationship to perceived rate and streaming. *Journal of Neurophysiology*, *97*, 2230–2238.
- Winer, J. A., Miller, L. M., Lee, C. C., & Schreiner, C. E. (2005). Auditory thalamocortical transformation: Structure and function. *Trends in Neurosciences*, *28*, 255–263.
- Winkler, I., Kushnerenko, E., Horvath, J., Ceponiene, R., Fellman, V., Huotilainen, M., et al. (2003). Newborn infants can organize the auditory world. *Proceedings of the National Academy of Sciences of the United States of America*, *100*, 11812–11815.
- Winkler, I., Sussman, E., Tervaniemi, M., Horvath, J., Ritter, W., & Näätänen, R. (2003). Preattentive auditory context effects. *Cognitive Affective and Behavioral Neuroscience*, *3*, 57–77.
- Winkler, I., Takegata, R., & Sussman, E. (2005). Event-related brain potentials reveal multiple stages in the perceptual organization of sound. *Cognitive Brain Research*, *25*, 291–299.
- Woldorff, M. G., Hillyard, S. A., Gallen, C. C., Hampson, S. R., & Bloom, F. E. (1998). Magnetoencephalographic recordings demonstrate attentional modulation of mismatch-related neural activity in human auditory cortex. *Psychophysiology*, *35*, 283–292.
- Yabe, H., Tervaniemi, M., Sinkkonen, J., Huotilainen, M., Ilmoniemi, R. J., & Näätänen, R. (1998). Temporal window of integration of auditory information in the human brain. *Psychophysiology*, *35*, 615–619.
- Yabe, H., Winkler, I., Czigler, I., Koyama, S., Kakigi, R., Sutoh, T., et al. (2001). Organizing sound sequences in the human brain: The interplay of auditory streaming and temporal integration. *Brain Research*, *897*, 222–227.

Received November 9, 2006

Revision received March 26, 2007

Accepted April 9, 2007 ■

Call for Nominations

The Publications and Communications (P&C) Board of the American Psychological Association has opened nominations for the editorships of **Psychological Assessment**, **Journal of Family Psychology**, **Journal of Experimental Psychology: Animal Behavior Processes**, and **Journal of Personality and Social Psychology: Personality Processes and Individual Differences (PPID)**, for the years 2010–2015. Milton E. Strauss, PhD, Anne E. Kazak, PhD, Nicholas Mackintosh, PhD, and Charles S. Carver, PhD, respectively, are the incumbent editors.

Candidates should be members of APA and should be available to start receiving manuscripts in early 2009 to prepare for issues published in 2010. Please note that the P&C Board encourages participation by members of underrepresented groups in the publication process and would particularly welcome such nominees. Self-nominations are also encouraged.

Search chairs have been appointed as follows:

- **Psychological Assessment**, William C. Howell, PhD, and J Gilbert Benedict, PhD
- **Journal of Family Psychology**, Lillian Comas-Diaz, PhD, and Robert G. Frank, PhD
- **Journal of Experimental Psychology: Animal Behavior Processes**, Peter A. Ornstein, PhD, and Linda Porrino, PhD
- **Journal of Personality and Social Psychology: PPID**, David C. Funder, PhD, and Leah L. Light, PhD

Candidates should be nominated by accessing APA's EditorQuest site on the Web. Using your Web browser, go to <http://editorquest.apa.org>. On the Home menu on the left, find "Guests." Next, click on the link "Submit a Nomination," enter your nominee's information, and click "Submit."

Prepared statements of one page or less in support of a nominee can also be submitted by e-mail to Emnet Tesfaye, P&C Board Search Liaison, at etesfaye@apa.org.

Deadline for accepting nominations is **January 10, 2008**, when reviews will begin.