Neural correlates of auditory scene analysis and perception

Kate L. Christison-Lagay[#1], **Adam M. Gifford**[#1], and **Yale E. Cohen**[2,3,4]

[1]Neuroscience Graduate Group, Perelman School of Medicine, University of Pennsylvania, Philadelphia, 19104

[2]Department of Otorhinolaryngology, University of Pennsylvania, Philadelphia, 19104

[3]Neuroscience, Perelman School of Medicine, University of Pennsylvania, Philadelphia, 19104

[4]Department of Bioengineering University of Pennsylvania, Philadelphia, 19104

[#] These authors contributed equally to this work.

Abstract

The auditory system is designed to transform acoustic information from low-level sensory representations into perceptual representations. These perceptual representations are the computational result of the auditory system's ability to group and segregate spectral, spatial and temporal regularities in the acoustic environment into stable perceptual units (i.e., sounds or auditory objects). Current evidence suggests that the cortex--specifically, the ventral auditory pathway--is responsible for the computations most closely related to perceptual representations. Here, we discuss how the transformations along the ventral auditory pathway relate to auditory percepts, with special attention paid to the processing of vocalizations and categorization, and explore recent models of how these areas may carry out these computations.

Keywords

auditory object; auditory stream; auditory scene analysis; categorization; perception; cortex

1. Introduction

Imagine, for a moment, that you are at a cocktail party and surrounded by sounds: music plays in the background; your conversation partner is telling you a story; beyond you, a group of fellow party-goers are engaged in a lively debate; and a cell phone is ringing. Despite the fact that these sound sources (e.g., the phone and the stereo speaker) are happening in close temporal and spatial proximity, with many likely having similar frequency components, you are readily— and seemingly effortlessly—able to differentiate

Corresponding author: Yale E. Cohen, PhD Department of Otorhinolaryngology 3400 Spruce St – 5 Ravdin Philadelphia, PA 19104 ycohen@mail.med.upenn.edu Phone: 1 215 898 7504 Fax: 1 215 898 9994.

between these different sounds. But how is our auditory system able to transform the acoustic information that is a mixture of the stimuli produced by each sound source, into distinct perceptual representations (i.e., sounds, such as the music or the cell phone's ring)?

The study of this problem is often referred to as 'auditory scene analysis', referencing Bregman's seminal book (1990; van Noorden, 1975b). Sounds or "auditory objects" are thought to be formed as a result of the auditory system's ability to detect, extract, segregate, and group the spatial, spectral, and temporal regularities in the acoustic environment into distinct perceptual units (i.e., *auditory objects* or sounds) (Bregman, 1990; McDermott, 2010; Shinn-Cunningham, 2008; Sussman et al., 2005; Winkler et al., 2009). Moreover, because audition is inherently temporal, a sound can span multiple acoustic events that unfold over time and a sequence of auditory objects forms an *auditory stream* (Bizley and Cohen, 2013; Bregman, 1990; Fishman et al., 2004; Micheyl et al., 2007; Sussman et al., 2007a). Auditory objects and streams are the basis from which, through categorization, we can reason about and respond adaptively to the auditory environment. In the following review, we explore the conceptual framework for auditory perception and delve into the role of the cortex in mediating auditory-object and stream formation, as well as its role in auditory categorization.

2. Auditory scene analysis, streaming, and predictive regularity

According to Bregman's (1990) theory, auditory scene analysis occurs in two stages: (1) the formation of alternative organizational schemes for incoming acoustic information, and (2) the selection of one of the alternative organizational schemes to be perceived. A simple example of this two-stage process is the sound sequence displayed in Fig. 1a, which consists of tone bursts A and B, each having a different frequency. There are two ways this sequence is generally heard: (1) group all acoustic events into a single stream (Fig. 1b top) or (2) segregate the tones of different frequency into two separate streams (Fig. 1b bottom). In the first scheme, a listener would hear a galloping rhythm of tones. In the second scheme, a listener would hear two distinct streams: one consisting of tone sequences at tone A's frequency and one at tone B's frequency. Indeed, humans can hear this sequence as either one or two streams (Cusack, 2005; Denham and Winkler, 2006), suggesting that both organizational schemes can be formed.

When multiple organizational schemes are formed, one must be selected for perception. Studies using alternating tone sequences, or those featured in Fig. 1, suggest that initially sequences tend to be perceived as a single stream but that over time the percept tends to change to hearing two distinct streams (Bregman, 1990; Cusack et al., 2004; Micheyl et al., 2005). This suggests that early in scene analysis, the default organizational scheme is to tend to integrate all events into a single stream and only over time does evidence 'build up' in support of an alternative scheme that segregates events into multiple streams. Perhaps a more direct example of active competition between competing organizational schemes is that of a perceptually bistable stimulus, which can be heard as either one or two streams. The perception of this stimulus as either a single or multiple streams tends to fluctuate spontaneously over time (Denham and Winkler, 2006), which suggests that both organizational schemes are represented simultaneously and compete for perception.

These organizational schemes reflect the potential ways in which the brain can parse the acoustic environment into different auditory streams. Auditory streams are formed on the basis of detected patterns, or *regularities*, from within the acoustic environment, under the reasonable assumption that auditory stimuli produced from a single sound source are likely to be more similar than those produced by different sources. For example, a male's voice can be differentiated from a female's voice based on differences in the voices' formant structures. Indeed, a large literature of psychoacoustic studies has demonstrated that auditory stimuli can be segregated into unique streams based on distinct regularities in frequency (van Noorden, 1975a), spatial location (Hill et al., 2011), timbre (Singh, 1987), and others (Bregman, 1990; Grimault et al., 2002; Vliegen and Oxenham, 1999). Thus, organizational schemes reflect the detection of distinct regularities within the acoustic environment. Moreover, these detected regularities are also inherently predictive of future acoustic events that the putative sound sources are likely to produce. Thus, auditory objects and streams can be thought of as *predictive-regularity representations* (Denham and Winkler, 2006; Winkler et al., 2009). In this sense, alternative organizational schemes compete on the basis of how well they predict future acoustic events.

Neurophysiological evidence in support of such a predictive-regularity hypothesis has come from the numerous electroencephalographic (EEG) and magnetoencephalographic (MEG) studies on deviance detection (Winkler et al., 2009). In a typical experiment, a sequence of tones (e.g., a sequence of tone bursts at 1 kHz) or other auditory stimuli is presented in a regular pattern. This regularity is disrupted by an occasional stimulus that deviates from this pattern (e.g., a tone burst at 2 kHz). If the brain was creating predictive-regularity representations, then it is reasonable to hypothesize that the brain should differentially encode events that deviate from the expected pattern. Indeed, the 'mismatch negativity' (MMN) is a frontally negative event-related response that reflects the differential response to deviant versus non-deviant stimuli (Winkler et al., 2009). The MMN can be measured in response to various types of deviations, ranging from simple changes in the acoustic features of a tone sequence (Alain et al., 1999; Horvath et al., 2001; Kisley et al., 2004; Schröger et al., 1992), to more abstract deviations, such as changes in the temporal order of an acoustic pattern (Korzyukov et al., 2003; Tervaniemi et al, 1994) and omissions (Yabe et al., 1997; Yabe et al., 2001). Critically, the MMN is elicited on the basis of the perceived organization of the stimulus (Ritter et al., 2000; Sussman et al., 2007b; Winkler et al., 2005; Yabe et al., 2001), which suggests that it reflects processes that follow the selection of a predictive-regularity representation and may signal the need for the auditory system to adjust its representation of the auditory environment (Winkler et al., 2009).

Current-source density analyses have distinguished two distinct neuronal generators of the MMN signal: one located within auditory cortex along the superior temporal gyrus (Giard et al., 1990; Hari et al., 1984; Molholm et al., 2005), and another located in frontal cortices (Doeller et al., 2003; Giard et al., 1990; Lappe et al., 2013; Molholm et al., 2005). These generators are also distinguished in their proposed functions, with temporal generators largely reflecting the detection of feature-specific deviations (Giard et al., 1990; Hari et al., 1984; Molholm et al., 2005) and frontal generators potentially reflecting involuntary

switching of attention in response to the changes in the acoustic environment (Giard et al., 1990; Molholm et al., 2005; Näätänen and Alho, 1995).

Studies on the underlying cortical mechanisms of the mismatch signal have found MMN-like activity in the core auditory field A1 (Fishman and Steinschneider, 2012; Javitt et al., 1994), suggesting that the detection of deviations from predictive regularities begin early on in cortical processing. However these signals may be more reflective of the effects of stimulus-specific adaptation (SSA) as opposed to a distinct and general deviance detection signal *per se* (Fishman, 2013; Fishman and Steinschneider, 2012). Since the MMN signal is distributed over a complex temporo-frontal cortical network (Garrido et al., 2009; Korzyukov et al., 2003; Näätänen et al., 2007; Schönwiesner et al., 2007; Szalárdy et al., 2013), it is possible that more general deviance detection signals that represent specifically the identification of violations from predictive regularities are located further downstream along the auditory pathway. Currently, the interpretation of the MMN as a representation of a top-down, predictive regularity representation versus an obligatory response relating to SSA effects is under debate (see Fishman, 2013 for details).

3. Attention and auditory scene analysis

The auditory system must not merely form these predictive regularity representations based upon the spectrotemporal regularities in an acoustic stimulus, but it must do so in the presence of competing acoustic information from other sound sources. This requires that the auditory system can identify and distinguish between the spectrotemporal regularities that are characteristic of distinct sound sources and parse the acoustic scene appropriately. Moreover, the auditory system must be able to track the evolution of these distinct regularities over time in order to maintain a Mable perceptual representation of a particular stream in the presence of these competing sounds. To accomplish this, the auditory system is capable of selectively attending to a particular stream and ignoring the competing sound sources.

Selective attention is the process by which the brain biases the processing of a particular object of interest at the expense of other objects in the environment. Recent electrophysiological work suggests that selective attention acts by enhancing the cortical representations of the attended sound streams, with population-level activity in the presence of competing sound streams reflecting primarily the activity in response to the attended stream in isolation compared to that of the ignored stream (Golumbic et al., 2013; Horton et al., 2013; Kerlin et al., 2010). These enhancements have been found primarily in low-frequency oscillatory phase (Golumbic et al., 2013; Kerlin et al., 2010) as well as high-frequency oscillatory power (Mesgarani and Chang, 2012a). Additionally, the auditory system seems to also suppress activity related to the competing stream (Horton et al., 2013), further enhancing the relative attentional gain on the attended stream. This attentional gain has also been shown to be frequency-specific (Da Costa et al., 2013; Lakatos et al., 2013), suggesting that selective attention may act as a feature-specific filter.

4. Representation of auditory objects

As noted above, auditory objects are reflections of the auditory system's transformation of an acoustic waveform from low-level sensory representations into perceptual representations. In the cortex, auditory information is hypothesized to be processed in two parallel pathways: a ventral pathway and a dorsal pathway (Griffiths, 2008; Kaas and Hackett, 1999; Rauschecker and Scott, 2009; Romanski et al., 1999). The ventral pathway is generally believed to mediate auditory perception by processing the content, identity, and meaning of a stimulus. The ventral auditory pathway begins in the core auditory fields A1 and R, which project to the anterolateral and middle-lateral belt regions of the auditory cortex (Kaas and Hackett, 2000; Rauschecker and Tian, 2000). These brain regions then project directly and indirectly to the ventrolateral prefrontal cortex (vlPFC) (Romanski et al., 1999). In contrast, the dorsal pathway mediates those computations underlying sound localization and the preparation of motor actions in response to those sounds, even in the absence of conscious perception or identification (Griffiths, 2008; Kaas and Hackett, 1999; Rauschecker and Scott, 2009; Romanski et al., 1999). In general, these computations fall under the heading of "audiomotor behaviors". Although the dorsal pathway clearly contributes to auditory perception (Rauschecker, 2012), our focus will be on the contributions of the ventral pathway.

Whether the neural processing underlying auditory perception is strictly hierarchical or distributed is still an area of active research, and there is evidence to support both positions. For example, the results of several studies support the existence of specific areas that encode pitch (Bendor and Wang, 2005; Bizley et al., 2013; Patterson et al., 2002; Penagos et al., 2004; Warren and Griffiths, 2003). On the other hand, there is equally compelling evidence to suggest that pitch-processing is served by a number of cortical areas throughout the ventral pathway (Bizley et al., 2010; Bizley et al., 2013; Bizley et al., 2009; Garcia et al., 2010; Griffiths et al., 2010; Hall and Plack, 2009 ; Staeren et al., 2009).

Moreover, although the degree to which a brain area represents specific sound features is a matter of debate, it is clear that different areas of the cortex tend to be sensitive to different sound features. In the core auditory fields, neural activity is sensitive to a number of sound features, such as a stimulus' frequency, intensity and location, as well as more derived properties, such as timbre and stimulus novelty (Bendor and Wang, 2005; Bizley and Walker, 2009; Bizley et al., 2010; Bizley et al., 2013; Bizley et al., 2009; Javitt et al, 1994; Razak, 2011; Schebesch et al., 2010 ; Ulanovsky et al., 2004; Versnel and Shamma, 1998 ; Wang et al., 1995; Watkins and Barbour, 2011; Werner-Reiss and Groh, 2008; Zhou and Wang, 2010). Neurons in the anterolateral belt (ALB) prefer more complex stimuli, such as band-passed noise, frequency-modulated sweeps, and vocalizations (Rauschecker and Tian, 2000, 2004; Rauschecker et al., 1995; Tian and Rauschecker, 2004; Tian et al., 2001). Other areas show a greater degree of stimulus selectivity, such as neurons in the anterior portions of the temporal lobe, which are highly selective for individual vocalizations, and the voice of specific vocalizers. Finally, activity in the ventrolateral prefrontal cortex (vlPFC) seems to represent a processing stage beyond that of sensory processing since vlPFC activity reflects non-spatial auditory attention, auditory working memory, sound meaning, and multimodal sensory integration (Cohen et al., 2009; Gifford III et al, 2005; Lee et al., 2009a;

Ng et al, 2013; Plakke et al, 2013a; Plakke et al., 2012; Plakke et al., 2013b; Poremba et al., 2004; Romanski et al., 2005; Romanski and Goldman-Rakic, 2002; Russ et al, 2008a; Russ et al., 2008b).

Whether neural activity reflects sensory components of a stimulus or the perceptual processing of that stimulus can be assessed directly by recording neural activity while animals are engaged in auditory tasks. For example, studies in the primary auditory cortex have found that neural activity correlates with behavioral reports of pitch (Bizley et al., 2010; Bizley et al., 2013; Bizley et al., 2009), amplitude modulation (Niwa et al., 2012), and tone contours (Selezneva et al., 2006). Although behaviorally correlated neural activity in the core does not necessarily mean the core is the locus of perceptual decisions (Gold and Shadlen, 2000, 2007), it does suggest that an important components of auditory-object processing occurs in the core.

Indeed, the results from several other studies suggest that neural correlates of perception are found in later portions of the ventral pathway. For example, MEG data suggest that the neural correlates of a listener hearing a sound, while engaged in an informational-masking paradigm, do not appear until the secondary (belt) auditory cortex (Gutschalk et al., 2008). Similarly, perceptual judgments of communication sounds (species-specific vocalizations and speech sounds) have also been found in belt region of the auditory cortex and higher auditory cortices (Binder et al., 2004; Chang et al., 2010; Christison-Lagay et al., 2011; Mesgarani and Chang, 2012b). Alternately, in studies of phonemic categorization, perceptual judgments are not found in the auditory cortex, but rather do not emerge until the level of the vlPFC (Lee et al., 2009a; Russ et al., 2008b; Tsunada et al., 2011).

It is unclear why some studies find choice-related activity as early as the core auditory cortex and others do not find it until much further downstream. However, it *is* clear that different parts of the auditory cortex have different stimulus preferences. Therefore, it is possible that the choice and complexity of stimuli and task may contribute to differences in where a stimulus is first perceived.

5. Representation of auditory streams

The core auditory cortex may play an important role in the neural encoding of streaming. In response to alternating ABAB tone sequences (see Fig. 1), A1 responses adapt to repeated presentations of a tone, as a functioning frequency and repetition rate. Specifically, a neuron increasingly adapts (i.e., fires less) as the repetition rate of the tones increases and as the spectral distance between the tone's frequency and the neuron's preferred frequency increases (Fishman et al., 2004; Fishman et al., 2001; Micheyl et al., 2005). In other words, A1 neurons respond more to their preferred frequency and are less suppressed by repeated presentations of this frequency than this "non-preferred" frequency.

Since A1 is tonotopically organized (Eggermont, 2001; Steinschneider et al., 1990), this pattern of firing rate suggests that the neural bases for stream segregation may be a place code. Under this hypothesis, the number of perceived streams is a function of the number of spatially separable active neural populations. In other words, this differential suppression acts to spatially separate the active neural populations such that only those neurons that are

most sensitive to the tone frequency are active. The degree to which these populations can be discriminated is related directly to the level of differential suppression, as exemplified in Fig 2. For large frequency differences (or faster rates of repetition; Fig. 2a first panel), the degree of suppression is large (Fig. 2a second panel). Therefore, the populations responding to the different tones are minimally overlapping (Fig. 2a third panel) and could then be read out by downstream neurons as two distinct neural populations, giving rise to the percept of two streams. In contrast, for small frequency differences (or slower rates of repetition), differential suppression is minimal, leading to significant population overlap and a downstream readout of a single stream (Fig. 2b top three panels). In support of a neural-place code theory, a simple model of stream segregation based upon differential firing rates of neural populations in response to alternating tone sequences is consistent with the buildup of streaming in humans (Micheyl et al., 2005).

While we explored in detail some of streaming work with respect to acoustic frequency and rates of repetition, this neural-place code is not limited to situations where streaming occurs on the basis of tone freauency. Indeed, recent work in the cat supports a neural-place code for stream segregation based upon location differences, with neurons preferentially responding to a particular spatial location aggregating in regions of primary auditory cortex (Middlebrooks and Bremen, 2013). It is also important to note that evidence for neural-place codes exists in the auditory pathway much earlier than the cortex (Pressnitzer et al., 2008). Thus, it is clear that the cortex is not necessarily responsible for the separation of neural populations under particular conditions. Indeed, subcortical processing may even extend to early specialization in the processing of vocalizations (Owren and Rendall, 2001; Portfors et al., 2009; Rendall, 2009). That a neural place code is maintained throughout the core auditory cortex suggests that it plays a role in the subsequent perception of auditory streams. However, although subcortical processing certainly contributes to auditory perception, there is no evidence to suggest that it is sufficient for auditory perception. Therefore, here, we will limit our discussion to the contribution of the cortex to perception.

In addition to a neural-place code, more recent studies suggest that the timing of neural activity also plays an important role auditory streaming. Elhilali et al. (2009) found that when the tone bursts are presented synchronously—instead of asynchronously as described above, listeners report hearing one stream. However, the neural responses in A1 to these synchronous tone sequences were similar to those elicited by the (asynchronous) alternating tone sequences. According to the neural-place code hypothesis, synchronous tone sequences with large frequency separations should be perceived similarly to alternating tone sequences (compare population profiles in Fig. 2a and 2c), which was not the case (Elhilali et al., 2009).

To reconcile this paradox, Elhilali et al. (2009) proposed a temporal-coherence model of stream segregation: streams are formed on the basis of the detection of neural populations with temporally coherent activity. Thus, for synchronous tone sequences or alternating sequences with small frequency separations, the active neural population(s) would respond in a temporally coherent manner, which could be read out downstream as evidence for a single stream. On the other hand, tone sequences with large frequency separations produce

two neural populations responding in an anti-coherent manner and would be interpreted as a distinct auditory stream (Fig. 2a-c, bottom panels).

An argument for integrated temporal and place codes comes from work on environmental or background noise. While the response to environmental sounds at a population level is similar to the sum of activity in response to the different frequencies within the environmental sound, there are significant differences between the response to the environmental sound and individual stimuli at those frequencies (Rotman et al., 2001). This suggests that a simple place code is insufficient. Work from (Chandrasekaran et al., 2010) suggests that both spiking and LFP activity in response to environmental sounds are stimulus-locked, and that neural representation of environmental sounds is highly distributed across neural populations that use both place and temporal codes.

In the end, it is likely that both neural topography and temporal coherence play complementary roles in stream formation. While neural-place codes can explain the build-up of streaming, there is no clear explanation for the perceptual bi-stability of certain stimuli (Denham and Winkler, 2006). Alternatively, a strict interpretation of temporal coherence is also likely insufficient, as recent studies have found that temporally coherent sounds can, in fact, be streamed under certain conditions (Micheyl et al., 2013a; Micheyl et al., 2010; Micheyl et al., 2013b).

6. Potential mechanisms of temporal coherence

The temporal coherence model reflects an ever-increasing scientific interest in testing the relationship between auditory processing and neural-oscillatory activity (Giraud and Poeppel, 2012; Schroeder and Lakatos, 2009). Oscillatory activity represents the large-scale coordinated activity of neural populations over relatively long timescales, such as that needed for the temporal-coherence model of streaming. To be a putative mechanism for mediating temporal coherence processing, oscillatory activity should (1) reflect spectrotemporal regularities of acoustic sequences mat are useful in mediating streaming and (2) be related to the processing of acoustic events in a manner consistent with known phenomena of streaming.

Indeed, neural oscillations reliably entrain to tone sequences, with the frequency of the entrained oscillation corresponding to the repetition rate of the sequence (Lakatos, 2005; Lakatos et al., 2013). Additionally, oscillatory entrainment has been exhibited for patterned spectral modulations (Henry and Obleser, 2012; Luo et al., 2006; Patel and Balaban, 2000). Particularly, the phase-alignment of the entrained neural oscillation is directly related to the spectral distance between a cortical site's preferred frequency and the frequency of a tone burst: oscillations align to a high excitability phase during the presentation of a tone burst at the site's preferred-frequency but align to a low-excitability phase when the frequency is not preferred (Lakatos et al., 2013). Thus, neural oscillations could act as a spectrotemporal filter by differentially modulating the amount of a site's activity. In this manner, only the activity from sites that are sufficiently sensitive to a particular stimulus would be outputted for further downstream processing.

Despite these promising findings, it is still possible that oscillatory activity is simply epiphenomenal and not directly functionally relevant for auditory streaming (Shadlen and Movshon, 1999; Shadlen and Newsome, 1998). For instance, it is known that single neurons can act as coincidence detectors (Yin and Chan 1990) and information integrators (Huk and Shadlen, 2005), suggesting that computations relating to temporal coherence could occur without requiring a role for oscillatory activity. Coincidence detectors in relation to the processing of temporal coherence in auditory streaming have yet to be discovered, however, and therefore the exact mechanisms for mediating temporal coherence are not fully understood.

7. Auditory categorization

It is clear that auditory objects and stream formation are critical aspects of auditory perception, providing valuable information regarding the putative sound sources in the environment. However, an additional level of grouping of auditory information is also important: categorization. While objects and streaming are grouped on the basis of acoustics, categorization can be thought of as a grouping of objects or streams. Categorization is an adaptive process that provides high-level, abstracted representations of sensory information, which allows us to mentally manipulate, reason about, and respond adaptively to objects in the environment. Thus, in order to fully understand auditory perception, we must understand the processes involved in auditory categorization.

Similar to auditory-object formation, there is evidence to suggest that areas as early as the core auditory cortex may be have categorization-related activity. Specifically, the spatial pattern activation of the core auditory fields changes over the course of category learning: after categories are learned, stimuli within the same category evoke one spatial pattern of activity, whereas those in a different category evoke a second pattern (Ohl et al., 2001). Further supporting the role of core auditory regions in categorization is work from Selezneva et al. (2006), who found neural correlates of categorization during a contour-discrimination task. This is further supported by the work of Ley et al. (Ley et al., 2012), who trained participants in task specifically designed to create new category boundaries in an auditory task. Using fMRI, Ley and colleagues found that changes associated in processing these categories were seen in primary auditory cortex (as well as adjacent regions). However, it is also possible that the effects seen in core auditory regions may be due to feedback from higher areas (Buffalo et al., 2010; Fritz et al., 2010).

Other studies suggest that categorization-related activity occurs beyond the core fields. For example, when asked to categorize the human words *bad, dad*, and morphs of these words, ALB neurons in rhesus monkeys responded in a categorical fashion that mirrored their behavioral responses (Lee et al., 2009b; Tsunada et al., 2011). However, it is not until the level of the vlPFC that neural activity becomes correlated with the monkeys' choices during this categorization task (Lee et al., 2009a; Russ et al., 2008b). The encoding of conspecific vocalizations also supports a role for secondary and higher-order auditory cortex in categorization: voice-specific (categorizing across calls) and call-specific (categorizing across vocalizers) neurons are found in the anterior temporal lobe (Perrodin et al., 2011; Petkov et al., 2008). Evidence for special processing of speech-sound categories in both

anterior and posterior auditory cortex also comes from human fMRI research (Chang et al., 2010; Chevillet et al., 2013; Leaver and Rauschecker, 2010; Obleser et al., 2006; Obleser and Eisner, 2009; Obleser et al., 2010). Furthermore, categorical activity in these areas may be selective for human speech (and music) over other categories, such as birdsong and animal vocalizations (Leaver and Rauschecker, 2010). vlPFC neurons have also been shown to elicit category-related neural activity (Gifford III et al., 2005) since they code the "meaning" of vocalizations as opposed to their acoustic structure.

It is also important to note, however, that there is evidence for distributed representations of both low-level stimulus features and abstracted features or categories of the stimulus through the auditory cortex (including the posterior auditory cortex) (Belin et al., 2002; Bizley and Walker, 2009; Giordano et al., 2012; Stoereff et al., 2009; Zatorre et al., 2002). Overall, these studies suggest that while certain aspects of auditory perception are evident early on in the cortical pathway, a complete view of auditory processing requires contributions from a distributed cortical network.

8. Conclusions

In spite of the progress that has been made in the study of the neural mechanisms involved in auditory streaming, many questions yet remain as to how this process is accomplished. For instance, all of the previous neurophysiology studies have employed either passive listening conditions (Fishman et al., 2004; Fishman et al., 2001), or active listening conditions in tasks unrelated to auditory streaming (Lakatos et al., 2013; Micheyl et al., 2005) *per se*. A full understanding of the mechanisms responsible for auditory streaming will necessarily require a direct comparison of neural activity and behavioral reports of percepts. Moreover, it will become increasingly important to record large populations of neurons simultaneously over multiple cortical fields in order to accurately assess the roles of temporal coherence and place-codes in mediating auditory streaming, as well as further elucidating the specific roles of primary, non-primary, and multisensory cortical areas in stream formation and maintenance. These future studies across cortical areas should also address the degree to which each of these areas reflects the conscious percept of streams or the pre-conscious processing. Multiple types of stimuli should be used to establish whether streaming is a uniform, general process, or whether the complexity of the stream affects the manner in which it is processed. Furthermore, because categorization can be thought of as another form of auditory grouping and segregation, further study will need to be done to explore if and how activity associated with categorization affects streaming. Specifically, it will be important to determine the degree to which differential processing along the ventral pathway is dependent upon the acoustics of the stimuli, experience, or task demands.

Acknowledgements

References

Alain C, Cortese F, Picton TW. Event-related brain activity associated with auditory pattern processing. Neuroreport. 1999; 10:2429–2434. [PubMed: 10439476]

Belin P, Zatorre RJ, Ahad P. Human temporal-lobe response to vocal sounds. Cognitive Brain Research. 2002; 13:17–26. [PubMed: 11867247]

Bendor D, Wang X. The neuronal representation of pitch in primate auditory cortex. Nature. 2005; 436:1161–1165. [PubMed: 16121182]

Binder JR, Liebenthal E, Possing ET, Medler DA, Ward BD. Neural correlates of sensory and decision processes in auditory object identification. Nature neuroscience. 2004; 7:295–301.

Bizley JK, Cohen YE. The what, where and how of auditory-object perception. Nature Reviews Neuroscience. 2013; 14:693–707.

Bizley JK, Walker KM. Distributed sensitivity to conspecific vocalizations and implications for the auditory dual stream hypothesis. The Journal of neuroscience : the official journal of the Society for Neuroscience. 2009; 29:3011–3013. [PubMed: 19279236]

Bizley JK, Walker KM, King AJ, Schnupp JW. Neural ensemble codes for stimulus periodicity in auditory cortex. J Neurosci. 2010; 30:5078–5091. [PubMed: 20371828]

Bizley JK, Walker KM, Nodal FR, King AJ, Schnupp JW. Auditory cortex represents both pitch judgments and the corresponding acoustic cues. Current biology : CB. 2013; 23:620–625. [PubMed: 23523247]

Bizley JK, Walker KM, Silverman BW, King AJ, Schnupp JW. Interdependent encoding of pitch, timbre, and spatial location in auditory cortex. J. Neurosci. 2009; 29:2064–2075. [PubMed: 19228960]

Bregman, A. Auditory Scene Analysis: The Perceptual Organization of Sound. MIT Press; Cambridge, MA.: 1990.

Buffalo EA, Fries P, Landman R, Liang LL, Desimone R. A backward progression of attentional effects in the ventral stream. Proceedings of the National Academy of Sciences of the United States of America. 2010:361–365. [PubMed: 20007766]

Chandrasekaran C, Turesson HK, Brown CH, Ghazanfar AA. The influence of natural scene dynamics on auditory cortical activity. J Neurosci. 2010; 30:13919–13931. [PubMed: 20962214]

Chang EF, Rieger JW, Johnson K, Berger MS, Barbara NM, Knight RT. Categorical speech representation in human superior temporal gyrus. Nat Neurosci. 2010; 13:1428–1432. [PubMed: 20890293]

Chevillet MA, Jiang X, Rauschecker JP, Riesenhuber M. Automatic phoneme category selectivity in the dorsal auditory stream. J Neurosci. 2013; 33:5208–5215. [PubMed: 23516286]

Christison-Lagay, KL.; Bennur, S.; Cohen, YE. Annual Meeting for the Society for Neuroscience. Washington D.C.: 2011. Neural correlates of acoustic variability in conspecific vocalizations. In Press

Cohen YE, Russ BE, Davis SJ, Baker AE, Ackelson AL, Nitecki R. A functional role for the ventrolateral prefrontal cortex in non-spatial auditory cognition. Proc. Natl. Acad. Sci U.S.A. 2009; 106:20045–20050. [PubMed: 19897723]

Cusack R. The Intraparietal Sulcus and Perceptual Organization. J Cogn Neurosci. 2005; 17:641–651. [PubMed: 15829084]

Cusack R, Deeks J, Aikman G, Carlyon RP. Effects of Location, Frequency Region, and Time Course of Selective Attention on Auditory Scene Analysis. J Exp Psychol Hum Percept Perform. 2004; 30:643–656. [PubMed: 15301615]

Da Costa S, van der Zwaag W, Miller LM, Clarke S, Saenz M. Tuning In to Sound: Frequency-Selective Attentional Filter in Human Primary Auditory Cortex. J Neurosci. 2013:1858–1863. [PubMed: 23365225]

Denham SL, Winkler I. The role of predictive models in the formation of auditory streams. J Physiol Paris. 2006; 100:154–170. [PubMed: 17084600]

Doeller CF, Opitz B, Mecklinger A, Krick C, Reith W, Schroger E. Prefrontal cortex involvement in preattentive auditory deviance detection:. Neuroimage. 2003; 20:1270–1282. [PubMed: 14568496]

Eggermont JJ. Between sound and perception: reviewing the search for a neural code. Hear Res. 2001:1–42. [PubMed: 11470183]

Elhilali M, Ma L, Micheyl C, Oxenham AJ, Shamma SA. Temporal Coherence in the Perceptual Organization and Cortical Representation of Auditory Scenes. Neuron. 2009; 61:317–329. [PubMed: 19186172]

Fishman YI. The Mechanisms and Meaning of the Mismatch Negativity. Brain Top. 2013

Fishman YI, Arezzo JC, Steinschneider M. Auditory stream segregation in monkey auditory cortex: effects of frequency separation, presentation rate, and tone duration. J Acoust Soc Am. 2004; 116:1656. [PubMed: 15478432]

Fishman YI, Reser DH, Arezzo JC, Steinschneider M. Neural correlates of auditory stream segregation in primary auditory cortex of the awake monkey. Hear Res. 2001; 151:167–187. [PubMed: 11124464]

Fishman YI, Steinschneider M. Searching for the Mismatch Negativity in Primary Auditory Cortex of the Awake Monkey: Deviance Detection or Stimulus Specific Adaptation? J Neurosci. 2012; 32:15747–15758. [PubMed: 23136414]

Fritz JB, David SV, Radtke-Schuller S, Yin P, Shamma SA. Adaptive, behaviorally gated, persistent encoding of task-relevant auditory information in ferret frontal cortex. Nature neuroscience. 2010; 13:1011–1019.

Garcia D, Hall DA, Plack CJ. The effect of stimulus context on pitch representations in the human auditory cortex. NeuroImage. 2010; 51:808–816.

Garrido MI, Kilner JM, Stephan KE, Friston KJ. The mismatch negativity: A review of underlying mechanisms. Clin Neurophysiol. 2009; 120:453–463. [PubMed: 19181570]

Giard M-H, Perrin F, Pernier J, Bouchet P. Brain generators implicated in the processing of auditory stimulus deviance: a topographic event-related potential study. Psychophysiol. 1990; 27:627–640.

Gifford GW III, MacLean KA, Hauser MD, Cohen YE. The neurophysiology of functionally meaningful categories: macaque ventrolateral prefrontal cortex plays a critical role in spontaneous categorization of species-specific vocalizations. Journal of Cognitive Neuroscience. 2005; 17:1471–1482. [PubMed: 16197700]

Giordano BL, McAdams S, Zatorre RJ, Kriegeskorte N, Belin P. Abstract Encoding of Auditory Objects in Cortical Activity Patterns. Cerebral cortex. 2012

Giraud A-L, Poeppel D. Cortical oscillations and speech processing: emerging computational principles and operations. Nat Neurosci. 2012; 15:511–517. [PubMed: 22426255]

Gold JI, Shadlen MN. Representation of a perceptual decision in developing oculomotor commands. Nature. 2000; 404:390–394. [PubMed: 10746726]

Gold JI, Shadlen MN. The neural basis of decision making. Ann. Rev. Neurosci. 2007; 30:535–574. [PubMed: 17600525]

Golumbic EMZ, Ding N, Bickel S, Lakatos P, Schevon CA, McKhann GM, Goodman RR, Emerson R, Mehta AD, Simon JZ, Poeppel D, Schroeder CE. Mechanisms Underlying Selective Neuronal Tracking of Attended Speech at a Cocktail Party. Neuron. 2013; 77:980–991. [PubMed: 23473326]

Griffiths TD. Sensory Systems: Auditory Action Streams? Curr. Biol. 2008; 18:R387–R388. [PubMed: 18460320]

Griffiths TD, Kumar S, Sedley W, Nourski KV, Kawasaki H, Oya H, Patterson RD, Brugge JF, Howard MA. Direct recordings of pitch responses from human auditory cortex. Current biology : CB. 2010; 20:1128–1132. [PubMed: 20605456]

Grimault N, Bacon SP, Micheyl C. Auditory stream segregation on the basis of amplitude-modulation rate. J Acoust Soc Am. 2002; 111:1340. [PubMed: 11931311]

Gutschalk A, Micheyl C, Oxenham AJ. Neural Correlates of Auditory Perceptual Awareness under Informational Masking. Plos Biol. 2008; 6:el38.

Hall DA, Plack CJ. Pitch processing sites in the human auditory brain. Cerebral cortex. 2009; 19:576–585. [PubMed: 18603609]

Hari R, Hämäläinen M, Ilmoniemi RJ, Kaukoranta E, Reinikainen K, Salminen J, Alho K, Näätänen R, Sams M. Responses of the primary auditory cortex to pitch changes in a sequence of tone pips: neuromagnetic recordings in man. Neurosci Lett. 1984; 50:127–132. [PubMed: 6493619]

Henry MJ, Obleser J. Frequency modulation entrains slow neural oscillations and optimizes human listening behavior. Proc Natl Acad Sci. 2012; 109:20095–20100. [PubMed: 23151506]

Hill KT, Bishop CW, Yadav D, Miller LM. Pattern of BOLD signal in auditory cortex relates acoustic response to perceptual streaming. BMC Neurosci. 12:85. [PubMed: 21849065]

Horton C, D'Zmura M, Srinivasan R. Suppression of competing speech through entrainment of cortical oscillations. J Neurophysiol. 2013; 109:3082–3093. [PubMed: 23515789]

Horváth J, Czigler I, Sussman ES, Winkler I. Simultaneously active pre-attentive representations of local and global rules for sound sequences in the human brain. Cogn Brain Res. 2001; 12:131–144.

Huk AC, Shadlen MN. Neural Activity in Macaque Parietal Cortex Reflects Temporal Integration of Visual Motion Signals during Perceptual Decision Making. J Neurosci. 2005; 25:10420–10436. [PubMed: 16280581]

Javitt DC, Steinschneider M, Schroeder CE, Vaughan HG Jr. Arezzo JC. Detection of stimulus deviance within primate primary auditory cortex: intracortical mechanisms of mismatch negativity (MMN) generation. Brain Res. 1994; 667:192–200. [PubMed: 7697356]

Kaas JH, Hackett TA. 'What' and 'where' processing in auditory cortex. Nat. Neurosci. 1999; 2:1045–1047. [PubMed: 10570476]

Kaas JH, Hackett TA. Subdivisions of auditory cortex and processing streams in primates. Proc Natl Acad Sci. 2000; 97:1179–11799.

Kerlin JR, Shahin AJ, Miller LM. Attentional Gain Control of Ongoing Cortical Speech Representations in a Cocktail Party. J Neurosci. 2010:620–628. [PubMed: 20071526]

Kisley MA, Davalos DB, Layton HS, Pratt D, Ellis JK, Seger CA. Small changes in temporal deviance modulate mismatch negativity amplitude in humans. Neurosci Lett. 2004; 358:197–200. [PubMed: 15039115]

Korzyukov OA, Winkler I, Gumenyuk VI, Alho K. Processing abstract auditory features in the human auditory cortex. Neuroimage. 2003; 20:2245–2258. [PubMed: 14683726]

Lakatos P. An Oscillatory Hierarchy Controlling Neuronal Excitability and Stimulus Processing in the Auditory Cortex. J Neurophysiol. 2005; 94:1904–1911. [PubMed: 15901760]

Lakatos P, Musacchia G, 0'Connel MN, Falchier AY, Javitt DC, Schroeder CE. The Spectrotemporal Filter Mechanism of Auditory Selective Attention. Neuron. 2013; 77:750761.

Lappe C, Steinstrater O, Pantev C. A Beamformer Analysis of MEG Data Reveals Frontal Generators of the Musically Elicited Mismatch Negativity. PLoS ONE. 2013; 8:e61296. [PubMed: 23585888]

Leaver AM, Rauschecker JP. Cortical representation of natural complex sounds: effects of acoustic features and auditory object category. J Neurosci. 2010; 30:7604–7612. [PubMed: 20519535]

Lee JH, Russ BE, Orr LE, Cohen Y. Prefrontal activity predicts monkeys' decisions during an auditory category task. Frontiers in Integrative Neuroscience. 2009a; 3

Lee, JH.; Tsunada, J.; Cohen, YE. Neural activity in the superior temporal gyrus during a discrimination task reflects stimulus category. Society for Neuroscience; Chicago, IL.: 2009b.

Ley A, Vroomen J, Hausfeld L, Valente G, De Weerd P, Formisano E. Learning of new sound categories shapes neural response patterns in human auditory cortex. J Neurosci. 2012; 32:13273–13280. [PubMed: 22993443]

Luo H, Wang Y, Poeppel D, Simon JZ. Concurrent Encoding of Frequency and Amplitude Modulation in Human Auditory Cortex: MEG Evidence. J Neurophysiol. 2006; 96:27122723.

McDermott J. The cocktail party problem. Curr Biol. 2010; 19:R1024–R1027. [PubMed: 19948136]

Mesgarani N, Chang EF. Selective cortical representation of attended speaker in multi-talker speech perception. Nature. 2012a; 485:233–236. [PubMed: 22522927]

Mesgarani N, Chang EF. Selective cortical representation of attended speaker in multi-talker speech perception. Nature. 2012b; 485:233–236. [PubMed: 22522927]

Micheyl C, Carlyon RP, Gutschalk A, Melcher JR, Oxenham AJ, Rauschecker JP, Tian B, Courtenay Wilson E. The role of auditory cortex in the formation of auditory streams. Hear Res. 2007; 229:116–131. [PubMed: 17307315]

Micheyl C, Hanson C, Demany L, Shamma SA, Oxenham AJ. Auditory Stream Segregation for Alternating and Synchronous Tones. J Exp Psychol Hum Percept Perform. 2013a

Micheyl C, Hunter C, Oxenham AJ. Auditory stream segregation and the perception of across-frequency synchrony. J Exp Psychol Hum Percept Perform. 2010; 36:1029–1039. [PubMed: 20695716]

Micheyl C, Kreft H, Shamma SA, Oxenham AJ. Temporal coherence versus harmonicity in auditory stream formation. J Acoust Soc Am. 2013b; 133:ELI88.

Micheyl C, Tian B, Carlyon RP, Rauschecker JP. Perceptual Organization of Tone Sequences in the Auditory Cortex of Awake Macaques. Neuron. 2005; 48:139–148. [PubMed: 16202714]

Middlebrooks JC, Bremen P. Spatial stream segregation by auditory cortical neurons. J Neurosci. 2013; 33:10986–11001. [PubMed: 23825404]

Molholm S, Martinez A, Ritter W, Javitt DC, Foxe JJ. The Neural Circuitry of Pre-attentive Auditory Change-detection: An fMRI Study of Pitch and Duration Mismatch Negativity generators. Cerebral Cortex. 2005; 15:545–551. [PubMed: 15342438]

Näätänen R, Alho K. Mismatch negativity- a unique measure of sensory processing in audition. Int J Neurosci. 1995; 80:317–337. [PubMed: 7775056]

Näätänen R, Paavilainen P, Rinne T, Alho K. The mismatch negativity (MMN) in basic research of central auditory processing: A review. Clin Neurophysiol. 2007; 118:2544–2590. [PubMed: 17931964]

Ng CW, Plakke B, Poremba A. Neural correlates of auditory recognition memory in the primate dorsal temporal pole. J Neurophysiol. 2013

Niwa M, Johnson JS, O'Connor KN, Sutter ML. Activity related to perceptual judgment and action in primary auditory cortex. J NSmosci. 2012; 32:3193–3210.

Obleser J, Boecker H, Drzezga A, Haslinger B, Hennenlotter A, Roettinger M, Eulitz C, Rauschecker JP. Vowel sound extraction in anterior superior temporal cortex. Hum Brain Mapp. 2006; 27:562–571. [PubMed: 16281283]

Obleser J, Eisner F. Pre-lexical abstraction of speech in the auditory cortex. Trends Cogn. Sci. 2009; 13:14–19. [PubMed: 19070534]

Obleser J, Leaver AM, Vanmeter J, Rauschecker JP. Segregation of vowels and consonants in human auditory cortex: evidence for distributed hierarchical organization. Frontiers in psychology. 2010; 1:232. [PubMed: 21738513]

Ohl FW, Scheich H, Freeman WJ. Change in pattern of ongoing cortical activity with auditory category learning. Nature. 2001; 412:733–736. [PubMed: 11507640]

Owren MJ, Rendall D. Sound on the rebound: bringing form and function back to the forefront in understanding nonhuman primate vocal signaling. Evolutionary Anthropology: Issues, News, and Reviews. 2001; 10:58–71.

Patel A, Balaban E. Temporal patterns of human cortical activity reflect tone sequence structure. Nature. 2000; 404:80–84. [PubMed: 10716446]

Patterson RD, Uppenkamp S, Johnsrude IS, Griffiths TD. The processing of temporal pitch and melody information in auditory cortex. Neuron. 2002; 36:767–776. [PubMed: 12441063]

Penagos H, Melcher JR, Oxenham AJ. A neural representation of pitch salience in nonprimary human auditory cortex revealed with functional magnetic resonance imaging. J. Neurosci. 2004; 24:6810–6815. [PubMed: 15282286]

Perrodin C, Kayser C, Logothetis NK, Petkov CI. Voice Cells in the Primate Temporal Lobe. Curr Biol. 2011; 21:1408–1415. [PubMed: 21835625]

Petkov CI, Kayser C, Steudel T, Whittingstall K, Augath M, Logothetis NK. A voice region in the monkey brain. Nat Neurosci. 2008; 11:367–374. [PubMed: 18264095]

Plakke B, Diltz MD, Romanski LM. Coding of vocalizations by single neurons in ventrolateral prefrontal cortex. Hear Res. 2013a; 305:135–143. [PubMed: 23895874]

Plakke, B.; Hwang, J.; Diltz, M.; Romanski, LM. Inactivation of ventral prefrontal cortex impairs audiovisual working memory. Society of Neuroscience New Orleans; LA: 2012. p. 87810

Plakke B, Ng CW, Poremba A. Neural Correlates of Auditory Recognition Memory in Primate Lateral Prefrontal Cortex. Neuroscience. 2013b

Poremba A, Malloy M, Saunders RC, Carson RE, Herscovitch P, Mishkin M. Species-specific calls evoke asymmetric activity in the monkey's temporal poles. Nature. 2004; 427:448–451. [PubMed: 14749833]

Portfors CV, Roberts PD, Jonson K. Over-representation of species-specific vocalizations in the awake mouse inferior colliculus. Neuroscience. 2009; 162:486–500. [PubMed: 19401221]

Pressnitzer D, Sayles M, Micheyl C, Winter IM. Perceptual Organization of Sound Begins in the Auditory Periphery. Curr Biol. 2008; 18:1124–1128. [PubMed: 18656355]

Rauschecker JP. Ventral and dorsal streams in the evolution of speech and language. Frontiers in evolutionary neuroscience. 2012; 4:7. [PubMed: 22615693]

Rauschecker JP, Scott SK. Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. Nat. Neurosci. 2009; 12:718–724. [PubMed: 19471271]

Rauschecker JP, Tian B. Mechanisms and streams for processing of "what" and "where" in auditory cortex. Proc. Natl. Acad. Sci U.S.A. 2000; 97:11800–11806. [PubMed: 11050212]

Rauschecker JP, Tian B. Processing of band-passed noise in the lateral auditory belt cortex of the rhesus monkey. J. Neurophysiol. 2004; 91:2578–2589. [PubMed: 15136602]

Rauschecker JP, Tian B, Hauser M. Processing of complex sounds in the macaque nonprimary auditory cortex. Science. 1995; 268:111–114. [PubMed: 7701330]

Razak KA. Systematic representation of sound locations in the primary auditory cortex. J Neurosci. 2011; 31:13848–13859. [PubMed: 21957247]

Rendall D. Asymmetries in the individual distinctiveness and maternal recognition of infant contact calls and distress screams in baboons. J. Acoust. Soc. Am. 2009; 125:1792. [PubMed: 19275336]

Ritter W, Sussman ES, Molholm S. Evidence that the mismatch negativity system works on the basis of objects. Neuroreport. 2000; 11:61–63. [PubMed: 10683830]

Romanski LM, Averbeck BB, Diltz M. Neural representation of vocalizations in the primate ventrolateral prefrontal cortex. Journal of Neurophysiology. 2005; 93:734–747. [PubMed: 15371495]

Romanski LM, Goldman-Rakic PS. An auditory domain in primate prefrontal cortex. Nat. Neurosci. 2002; 5:1Y16. [PubMed: 11753407]

Romanski LM, Tian B, Fritz J, Mishkin M, Goldman-Rakic PS, Rauschecker JP. Dual streams of auditory afferents target multiple domains in the primate prefrontal cortex. Nat. Neurosci. 1999; 2:1131–1136. [PubMed: 10570492]

Rotman Y, Bar-Yosef O, Nelken I. Relating cluster and population responses to natural sounds and tonal stimuli in cat primary auditory cortex. Hear. Res. 2001; 152:110–127. [PubMed: 11223286]

Russ BE, Ackelson AL, Baker AE, Cohen YE. Coding of auditory-stimulus identity in the auditory non-spatial processing stream. J. Neurophysiol. 2008a; 99:87–95. [PubMed: 18003874]

Russ BE, Orr LE, Cohen YE. Prefrontal neurons predict choices during an auditory same-different task. Curr. Biol. 2008b; 18:1483–1488. [PubMed: 18818080]

Schebesch G, Lingner A, Firzlaff U, Wiegrebe L, Grothe B. Perception and neural representation of size-variant human vowels in the Mongolian gerbil (Meriones unguiculatus). Hearing research. 2010; 261:1–8. [PubMed: 20004713]

Schönwiesner M, Novitski N, Pakarinen S, Carlson S, Tervaniemi M, Naatanen R. Heschl's Gyrus, Posterior Superior Temporal Gyrus, and Mid-Ventrolateral Prefrontal Cortex Have Different Roles in the Detection of Acoustic Changes. Journal of Neurophysiology. 2007; 97:2075–2082. [PubMed: 17182905]

Schroeder CE, Lakatos P. Low-frequency neuronal oscillations as instruments of sensory selection. Trends Neurosci. 2009; 32:9–18. [PubMed: 19012975]

Schröger E, Näätänen R, Paavilainen P. Event-related potentials reveal how non-attended complex sound patterns are represented by the human brain. Neurosci Lett. 1992; 146:183–186. [PubMed: 1491786]

Selezneva E, Scheich H, Brosch M. Dual Time Scales for Categorical Decision Making in Auditory Cortex. Curr. Biol. 2006; 16:2428–2433. [PubMed: 17174917]

Shadlen MN, Movshon JA. Synchrony unbound: a critical evaluation of the temporal binding hypothesis. Neuron. 1999; 24:67–77. 111–125. [PubMed: 10677027]

Shadlen MN, Newsome WT. The variable discharge of cortical neurons: implications for connectivity, computation, and information coding. J. Neurosci. 1998; 18:3870–3896. [PubMed: 9570816]

Shinn-Cunningham BG. Object-based auditory and visual attention. Trends Cogn Sci. 2008; 12:182–186. [PubMed: 18396091]

Singh PG. Perceptual organization of compex-tone sequences: A tradeoff between pith and timbre? J Acoust Soc Am. 1987; 82:886–889. [PubMed: 3655122]

Staeren N, Renvall H, De Martino F, Goebel R, Formisano E. Sound categories are represented as distributed patterns in the human auditory cortex. Current biology : CB. 2009; 19:498502.

Steinschneider M, Arezzo JC, Vaughan J, Herbert G. Tonotopic features of speech-evoked activity in primate auditory cortex. Brain Res. 1990; 519:158–168. [PubMed: 2397403]

Sussman ES, Bregman AS, Wang WJ, Khan FJ. Attentional modulation of electrophysiological activity in auditory cortex for unattended sounds within multistream auditory environments. Cognitive Affective & Behavioral Neuroscience. 2005; 5:93–110.

Sussman ES, Horvath J, Winkler I, Orr M. The role of attention in the formation of auditory streams. Percept Psychophys. 2007a; 69:136–152. [PubMed: 17515223]

Sussman ES, Wong R, Horváth J, Winkler I, Wang W. The development of the perceptual organization of sound by frequency separation in 5-11-year-old children. Hear Res. 2007b; 225:117–127. [PubMed: 17300890]

Szalárdy O, Böhm TM, Bendixen A, Winkler I. Event-related potential correlates of sound organization: Early sensory and late cognitive effects. Biol Psychol. 2013; 93:97–104. [PubMed: 23384511]

Tervaniemi M, Saarinen J, Paavilainen P, Danilova N, Naatanen R. Temporal integration of auditory information in sensory memory as reflected by the mismatch negativity. Biol Psychol. 1994; 38:157–167. [PubMed: 7873700]

Tian B, Rauschecker JP. Processing of frequency-modulated sounds in the lateral auditory belt cortex of the rhesus monkey. J. Neurophysiol. 2004; 92:2993–3013. [PubMed: 15486426]

Tian B, Reser D, Durham A, Kustov A, Rauschecker JP. Functional specialization in rhesus monkey auditory cortex. Science. 2001; 292:290–293. [PubMed: 11303104]

Tsunada J, Lee JH, Cohen YE. Representation of speech categories in the primate auditory cortex. J Neurophysiol. 2011

Ulanovsky N, Las L, Farkas D, Nelken I. Multiple time scales of adaptation in auditory cortex neurons. J Neurosci. 2004; 24:10440–10453. [PubMed: 15548659]

van Noorden, L. Temporal coherence in the perception of tone sequences. Technical University Eindhoven; 1975a. p. 1-127.

van Noorden, L. Temporal coherence in the perception of tone sequences. Technical University Eindhoven; 1975b. p. 1-127.

Versnel H, Shamma SA. Spectral-ripple representation of steady-state vowels in primary auditory cortex. The Journal of the Acoustical Society of America. 1998; 103:2502–2514. [PubMed: 9604344]

Vliegen J, Oxenham AJ. Sequential stream segregation in the absence of spectral cues. J Acoust Soc Am. 1999; 105:339–346. [PubMed: 9921660]

Wang XQ, Merzenich MM, Beitel R, Schreiner CE. Representation of a species-specific vocalization in the primary auditory cortex of the common marmoset: Temporal and spectral characteristics. Journal of Neurophysiology. 1995; 74:2685–2706. [PubMed: 8747224]

Warren JD, Griffiths TD. Distinct mechanisms for processing spatial sequences and pitch sequences in the human auditory brain. J Neurosci. 2003; 23:5799–5804. [PubMed: 12843284]

Watkins PV, Barbour DL. Level-tuned neurons in primary auditory cortex adapt differently to loud versus soft sounds. Cerebral Cortex. 2011; 21:178–190. [PubMed: 20457692]

Werner-Reiss U, Groh JM. A rate code for sound azimuth in monkey auditory cortex: implications for human neuroimaging studies. J Neurosci. 2008; 28:3747–3758. [PubMed: 18385333]

Winkler I, Denham SL, Nelken I. Modeling the auditory scene: predictive regularity representations and perceptual objects. Trends Cogn Sci. 2009; 13:532–540. [PubMed: 19828357]

Winkler I, Takegata R, Sussman ES. Event-related brain potentials reveal multiple stages in the perceptual organization of sound. Cogn Brain Res. 2005; 25:291–299.

Yabe H, Tervaniemi M, Reinikainen K, Naatanen R. Temporal window of integration revealed by MMN to sound omission. Neuroreport. 1997; 8:1971–1974. [PubMed: 9223087]

Yabe H, Winkler I, Czigler I, Koyama S, Kakigi R, Sutoh T, Hiruma T, Kaneko S. Organizing sound sequences in the human brain: the interplay of auditory streaming and temporal integration. Brain Res. 2001; 897:222–227. 5. [PubMed: 11282382]

Yin TCT, Chan JCK. Int/rauyd time sensitivity in medial superior olive ofcat. J Neurophysiol. 1990; 64:465–488. [PubMed: 2213127]

Zatorre RJ, Belin P, Penhune VB. Structure and function of auditory cortex: music and speech. Trends Cojfi. Sji. 2002; 6:37–46.

Zhou Y, Wang XQ. Cortical processing of dynamic sound envelope transitions. J Neurosci. 2010; 30:16741–16754. [PubMed: 21148013]

Highlights

1. Auditory system transforms acoustic stimuli into perceptual units called sounds

2. The ventral auditory pathway plays an important role in auditory perception

3. Both spatial and temporal models of rate coding needed for auditory perception
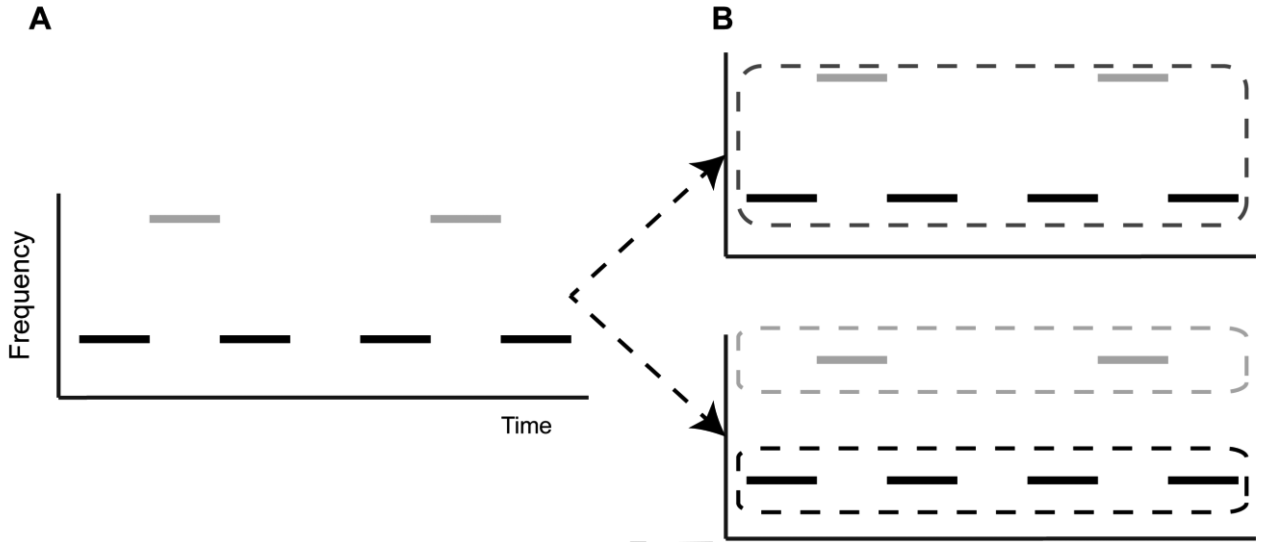
Figure 1.
Competition between alternative organizational schemes. Consider the sequence above sequence of tones (A), which consists of a repeating low-high-low pattern. (B) One potential organizational scheme could be to group all tones into a single group, which would be heard as a "gallop" (top). Alternatively, another scheme could be two segregate the tones of the same frequency into separate groups, which would lead to the perception of two isofrequency patterns (bottom). Depending on the timing and frequency separations between the tones, humans are capable of hearing both (Cusack, 2005; Denham and Winkler, 2006), consistent with the hypothesis that auditory perception functions on the basis of competing organizational schemes (Bregman, 1990).
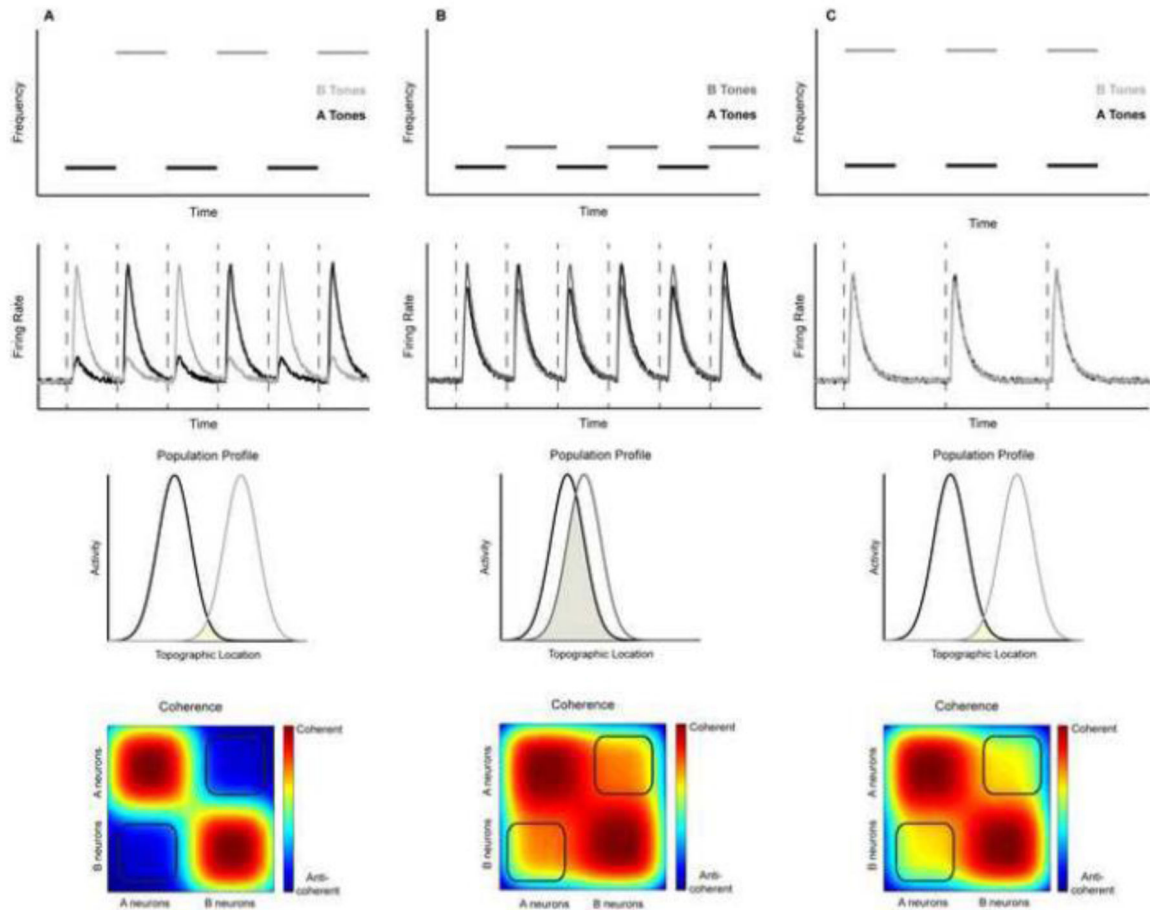
Figure 2.
Putative neural mechanisms mediating auditory streaming. The most extensive studying of the neural mechanisms of auditory streaming has focused on the processing of alternating tone sequences of varying frequency separations. (**A**) For alternating tone sequences with small frequency separations (first panel), neural populations that are most sensitive to one of the tones tend to be modulated by second tone as well (second panel). Given the tonotopic organization of A1 (Eggermont, 2001; Steinschneider et al., 1990), this would create essentially a single active population of neurons that could be read out downstream as representing a single stream (third panel). According to the temporal coherence model (Elhilali et al., 2009), the activity of the A and B populations exhibit high temporal coherence as well (fourth panel), which could also be read out downstream as a single stream. (**B**) For alternating tone sequences with large frequency separations (first panel), neural populations are only sensitive to the frequency of one of the tone bursts (second panel). Consequently, the A and B populations are effectively distinct populations (third panel) and can be read out downstream as two distinct streams. Consistent with this interpretation, the activity of the A and B populations is highly anti-coherent (fourth panel). (**C**) However, if the timing of the A and B tones is altered such that they are presented simultaneously (first panel), the story is slightly different with large frequency separations. Like with tone bursts with small frequency separations (see **A**), the neural populations are modulated by the frequencies of both tone bursts (second panel) but since the frequency

separation is large, there are two distinct populations of activity (third panel) like in **B**. This suggests that the simultaneous tones are activating largely separate neural populations and should be perceived as two distinct streams. However, unlike **B**, the activity in these populations is highly coherent and is heard as a singSystream (fourth panel). Figure was adapted from Elhilali et al., 2009 and Micheyl et al., 2005.