



Neurophysiological mechanisms involved in auditory perceptual organization

Aurelie Bidet-Caulet^{1*} and Olivier Bertrand^{2,3}

¹ Helen Wills Neuroscience Institute, University of California, Berkeley, CA, USA

² INSERM, U821, Lyon, France

³ University Lyon 1, Lyon, France

In our complex acoustic environment, we are confronted with a mixture of sounds produced by several simultaneous sources. However, we rarely perceive these sounds as incomprehensible noise. Our brain uses perceptual organization processes to independently follow the emission of each sound source over time. If the acoustic properties exploited in these processes are well-established, the neurophysiological mechanisms involved in auditory scene analysis remain unclear and have recently raised more interest. Here, we review the studies investigating these mechanisms using electrophysiological recordings from the cochlear nucleus to the auditory cortex, in animals and humans. Their findings reveal that basic mechanisms such as frequency selectivity, forward suppression and multi-second habituation shape the automatic brain responses to sounds in a way that can account for several important characteristics of perceptual organization of both simultaneous and successive sounds. One challenging question remains unresolved: how are the resulting activity patterns integrated to yield the corresponding conscious percepts?

Keywords: neurophysiology, electrophysiology, stream segregation, concurrent sound, auditory scene analysis

Edited by:

Leon Y. Deouell, The Hebrew University of Jerusalem, Israel;
Interdisciplinary Center for Neural Computation, The Hebrew University of Jerusalem, Israel

Reviewed by:

Daniel Pressnitzer, Ecole Normale Supérieure, France
Claude Alain, Rotman Research Institute, Canada

* Correspondence:



Aurelie Bidet-Caulet obtained her Ph.D. from the University Lyon 1, France, in 2007. During her thesis, she studied the brain mechanisms of human auditory perception in the laboratory of Dr. O. Bertrand, INSERM U821, Brain Dynamics and Cognition. She is currently a Post-doctoral Fellow in the laboratory of Pr. R.T. Knight at the University of California, Berkeley. Her current research focuses on the neurophysiological mechanisms of auditory attention. a.bidet-caulet@berkeley.edu

INTRODUCTION

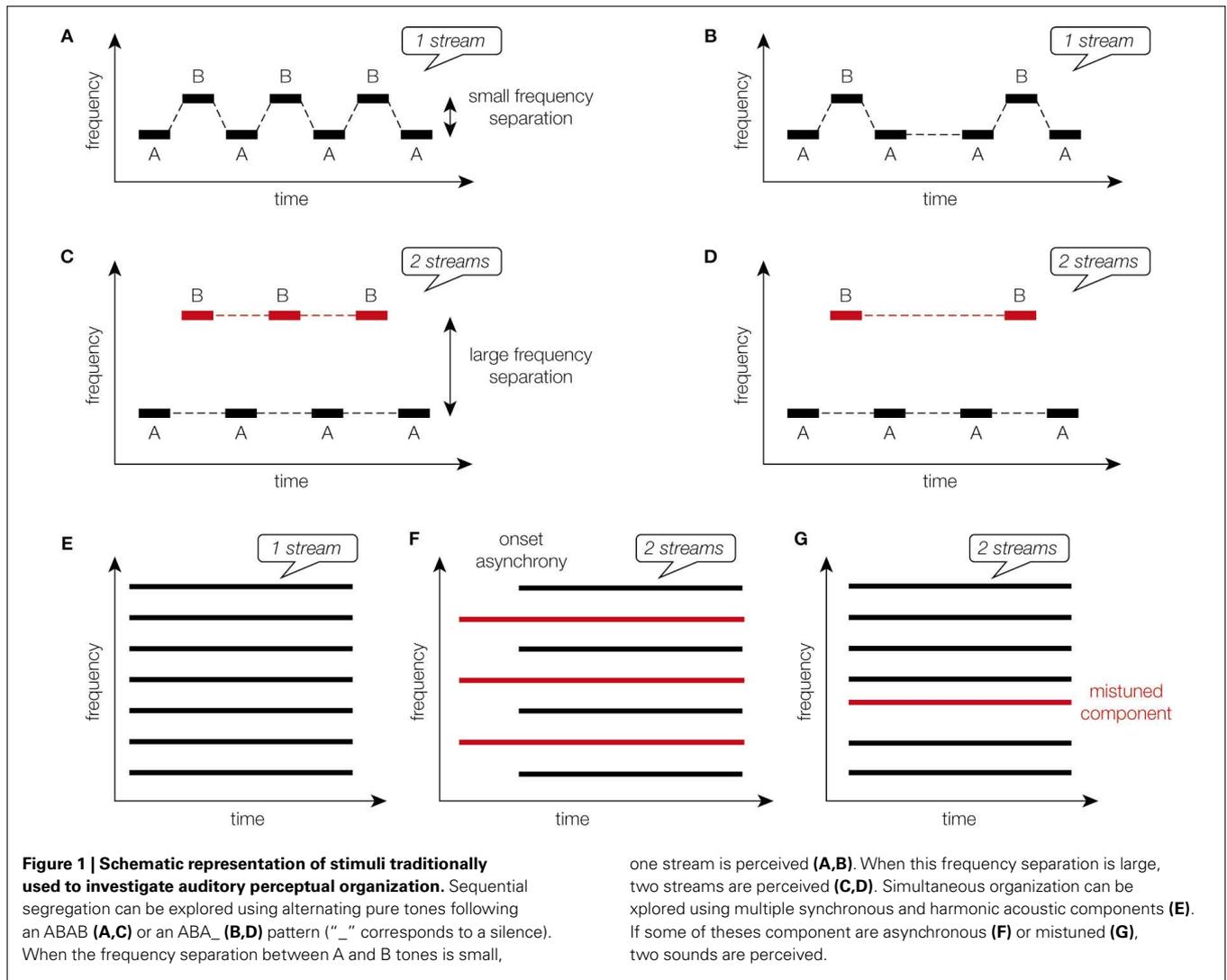
In our complex acoustic environment, we are confronted with a mixture of acoustic waves produced by several simultaneously active sources. However, we rarely perceive these sounds as incomprehensible noise. We are actually able to distinguish each **sound source** and to independently follow their emission over time. This capacity relies on the perceptual organization or auditory scene analysis: our auditory system groups together or dissociates **acoustic components** (or **events**) at a given instant and over time, leading to the perception of several simultaneous **auditory streams**.

PSYCHOLOGICAL FINDINGS ON AUDITORY SCENE ANALYSIS

The auditory scene analysis (ASA) has been largely investigated in psychoacoustical studies (reviewed in Bregman, 1990; Carlyon, 2004; Moore and

Gockel, 2002). According to these experiments, several acoustic properties can influence **sequential** or **simultaneous** perceptual organization of sounds, following Gestalt laws: similarity, proximity, common fate and good continuity.

Sequential organization has been mainly addressed with the “auditory streaming” phenomenon which corresponds to the perception of one or two streams from successive alternating acoustic events. Streaming is usually studied using simple sounds (**Figure 1**), such as sequences of pure tones alternating between two frequencies, A and B, according to a repeating ABAB or ABA_ABA pattern (‘_’ represents a silent gap). The percept induced by the tone sequence was found to depend on the frequency separation between the A and B tones and on the presentation rate (tempo) of these sounds. When the frequency difference is small and/or the tempo



is slow, the sequence is heard as one coherent stream of tones alternating in pitch. When the frequency difference is large and/or the tempo is fast, the sequence is perceived as two distinct streams, one with the frequency A and one with the frequency B. For intermediate separations and tempos, the percept can spontaneously flip from “one stream” to “two streams”, and *vice-versa*, with an increased probability to hear two streams as the sequence progresses (streaming buildup). Other parameters, such as intensity, spatial location or timbre, can also influence the perception of sound sequences if the separation is large enough.

Simultaneous organization has been far less explored and is highly influenced by the synchronization of onsets or offsets of overlapping acoustic events (Figure 1F). Acoustic events with an onset asynchrony inferior to 30 ms are more likely to be grouped into one auditory stream.

Harmonicity and pitch also play an important role in the simultaneous organization. Harmonic acoustic events share the same temporal periodicity and are usually heard as one sound, whereas a mistuned partial would rather be perceived as a pure tone separated from the other harmonic components (Figure 1G).

NEUROPHYSIOLOGICAL MECHANISMS UNDERLYING SIMULTANEOUS ORGANIZATION

Neurophysiological mechanisms involved in auditory simultaneous organization remain equivocal. The main reason is the difficulty to dissociate the neural activity specifically corresponding to either sound. This issue can be solved using overlapping sounds modulated at different frequencies, each sound eliciting a steady-state response at the same frequency as its amplitude

Sound source

Physical object emitting sounds.

Acoustic component or event

Sound emitted by a sound source.

Auditory stream

Mental representation of successive or overlapping acoustic events emitted by the same sound source.

Sequential organization

Segregation or integration of successive acoustic events over time into streams.

Simultaneous organization

Segregation or integration of overlapping acoustic events at a given instant into streams.

Frequency selectivity

Frequency selectivity refers to neurons in the auditory system that are responding to specific frequency band(s). In addition, at many levels of the auditory system, neurons are spatially organized so that neighboring neurons are selective to tones close to each other in frequency, resulting in a topographic organization called tonotopy.

Multi-second adaptation/habituation

Mechanism inducing a response decrement following stimulus repetition over hundreds of milliseconds and characterized by response recovery to a change stimulus and dishabituation to a previously habituated stimulus after a change stimulus.

modulation. In a recent study, the neurophysiological mechanisms involved in the perceptual organization of concurrent sounds were investigated using this electrophysiological tagging (21 and 29 Hz amplitude-modulated sounds) and intracortical signal directly recorded from the auditory cortex of epileptic patients (Bidet-Caulet et al., 2007A). This work revealed that different mechanisms, described in the following, are involved in the segregation or grouping of overlapping components as a function of the acoustic context.

FREQUENCY SELECTIVITY AND ONSET RESPONSE SYNCHRONIZATION

In this study, sound onset asynchrony was manipulated to induce the segregation or grouping of two concurrent components: either the 21-Hz component was starting 800 ms before the 29-Hz component (pitch continuity of the 21-Hz component), resulting in the perception of two streams; or the 21 and 29 Hz components were synchronous, preceded by a 21-Hz component at a distinct pitch (pitch discontinuity of the 21-Hz component), leading to the percept of one sound (Figure 2A). Transient evoked responses in secondary auditory areas (Figure 3) were found larger for pitch discontinuity than for pitch continuity. This can be explained by the **frequency selectivity** of auditory areas: the pitch discontinuity of the 21-Hz component activates a new neural population at the same time as the 29-Hz component onset, resulting in larger transient responses (Figures 2B,C). This finding suggests that synchronization of transient responses could account for grouping of overlapping auditory components.

FREQUENCY-SELECTIVE HABITUATION

During the overlap of the 21 and 29 Hz components (sound competition), the 21-Hz steady-state response (SSR), generated in the primary auditory cortex, PAC (Figure 3), was found larger for pitch discontinuity than for pitch continuity. A decrease of the 21-Hz SSR was observed over the course of the sound, suggesting the involvement of **habituation** mechanisms. In the case of onset asynchrony, 21-Hz SSR is continuously reduced by habituation mechanisms, resulting in a small 21-Hz SSR during sound competition; whereas in the case of onset synchrony, 21-Hz SSR reduction by habituation mechanisms is interrupted by the pitch discontinuity leading to a larger 21-Hz SSR during sound competition (Figure 2D). By varying the weight of the 21-Hz response (the 29-Hz SSR being unaffected), frequency-selective habituation mecha-

nisms modulate amplitude ratios between 29 and 21 Hz activities. In the case of pitch continuity, the 21-Hz response being highly reduced, the 29-Hz response becomes relatively more important. This could contribute to the increased saliency of the new-coming 29-Hz component, leading to the segregation into two streams. Conversely, in the case of pitch discontinuity, the 21-Hz response being slightly reduced, the 29-Hz component tends to merge into the acoustic mixture and the two components are grouped into one complex stream. Selective attention has been shown to modulate the cortical representation of concurrent sounds by increasing the SSR to relevant sound and decreasing it to irrelevant ones, resulting also in a modification of the amplitude ratio between the cortical representations of each sound (Bidet-Caulet et al., 2007B). One can imagine that when the ratio is largely in favor of one component, two distinct sounds would be perceived with one being more salient; whereas when the ratio is close to one, the components would be grouped and perceived as one complex sound. Thus, the interplay between habituation, attention and other mechanisms could influence the cortical representation of sounds and be involved in maintaining one percept and/or in percept shifting.

GAMMA OSCILLATORY ACTIVITIES AND PERCEPTUAL BINDING

In secondary auditory areas, induced oscillatory activities in the gamma range (50–90 Hz) were found more pronounced when one stream, rather than two, was perceived. This effect could be explained by the pitch discontinuity producing, in addition to the 29-Hz onset, a gamma response. However, oscillatory activities have been proposed as the neural mechanism promoting the interaction between different neural populations which are involved in processing distinct components of the same object (Tallon-Baudry and Bertrand, 1999), and thus could play an important role in the construction of coherent percepts (Vidal et al., 2006). Bidet-Caulet et al. (2007A) findings are consistent with this hypothesis, since induced gamma oscillations were larger when the two components are grouped together into one complex and coherent sound. Therefore, gamma oscillations could integrate and bind acoustic processing of the different components, which is performed by distinct groups of neurons, and directly reflect the auditory percept.

OTHER MECHANISMS

Concurrent sound perception has also been explored using scalp EEG and harmonic com-

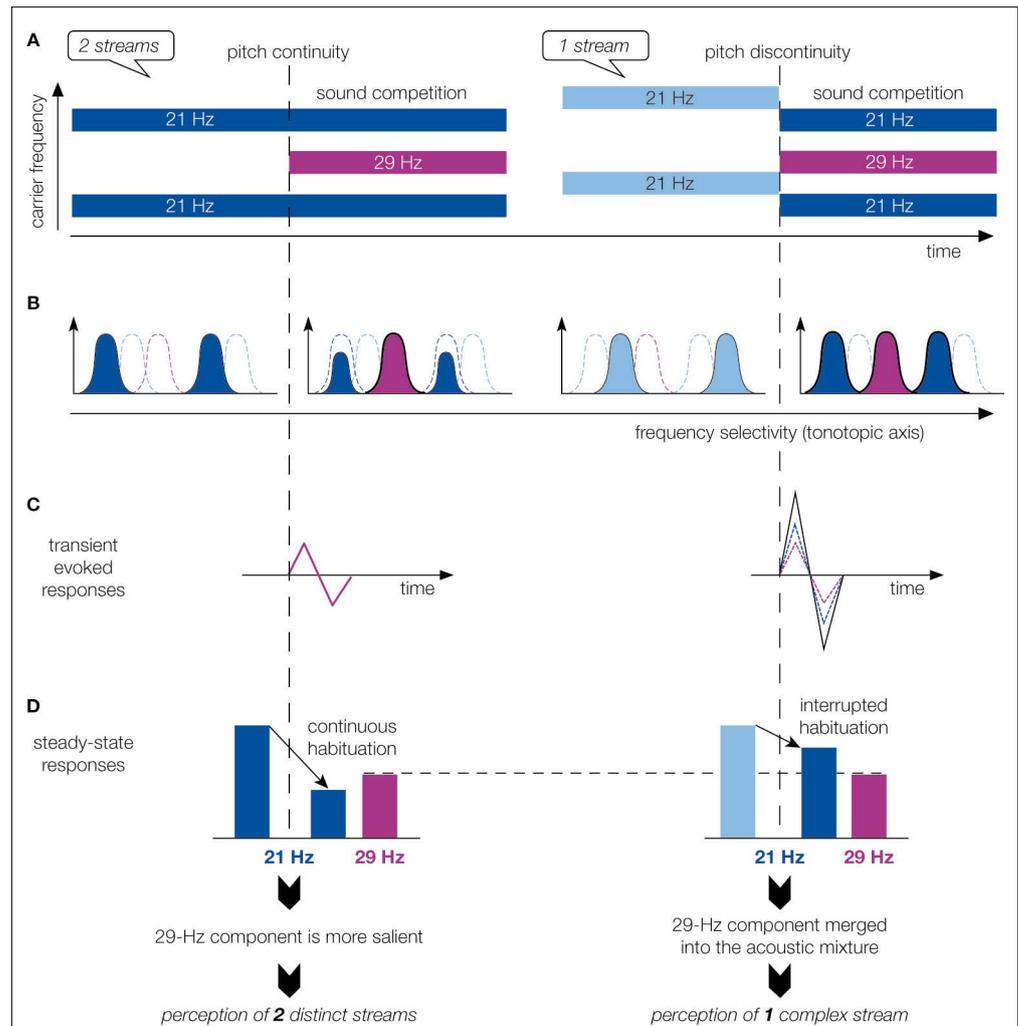
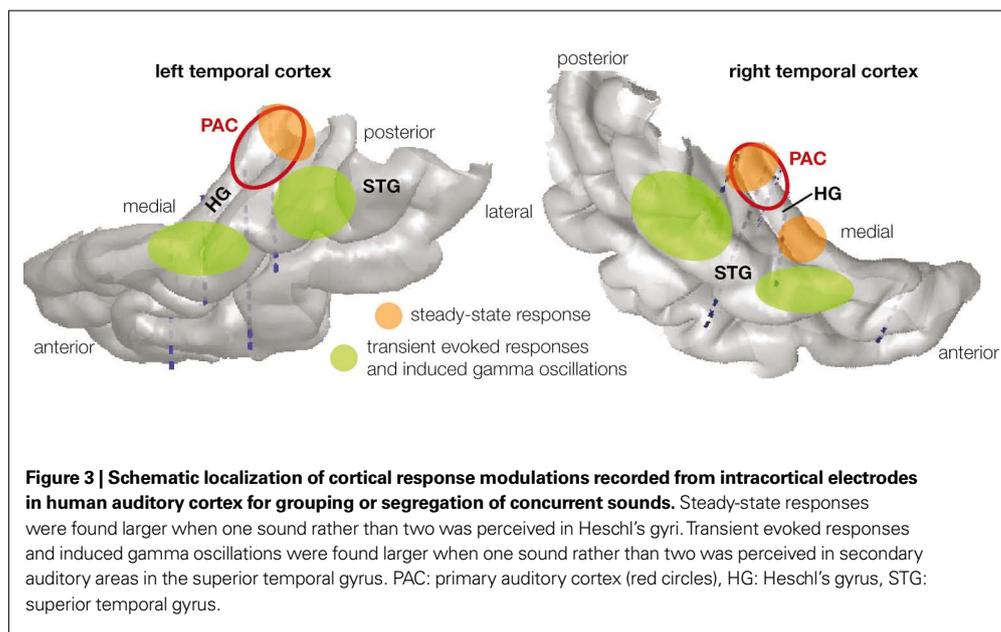


Figure 2 | Schematic representation and interpretation of response modulations recorded from intracortical electrodes in human auditory cortex during concurrent sound perception. (A) Stimuli. During sound competition, stimuli are identically composed of a 21 Hz amplitude-modulated component (blue) and a 29 Hz amplitude-modulated one (purple). When the sound competition period is preceded by the 21-Hz component at the same pitch (pitch continuity), 21-Hz and 29-Hz components are asynchronous and two streams are perceived (left panel). When the sound competition period is preceded by the 21-Hz component at a different pitch (pitch discontinuity), 21-Hz and 29-Hz components are synchronous and one stream is heard (right panel). **(B)** Schematic representation of the neural population activated by the stimuli in a tonotopically organized auditory area before and during sound competition. Bell-shape curves represent spatial activity patterns, along a tonotopic axis, evoked by the acoustic component represented in A with the same color. Bold curves indicate groups of neurons activated only during sound competition. Dashed curves indicate groups of neurons not activated during the considered period. **(C)** Schematic representation of the transient responses evoked at the onset of sound competition. In the case of pitch continuity (left panel), only responses to the 29-Hz component onset (purple) are observed. In the case of pitch discontinuity (right panel), responses to both the 29-Hz component onset (dashed purple) and the 21-Hz component pitch discontinuity (dashed blue) contribute to the larger recorded responses (black). **(D)** Schematic representation of the amplitude of steady-state response (SSR) at 21 and 29 Hz. 29-Hz SSR (purple) is observed only during sound competition with similar amplitudes in both conditions. 21-Hz SSR (blue) is observed during the whole stimulus duration, its amplitude is decreasing with time, probably because of habituation mechanisms. In the case of pitch continuity, habituation mechanisms are not interrupted, resulting in a 21-Hz SSR of small amplitude during sound competition. When habituation mechanisms are interrupted by the pitch discontinuity, the 21-Hz SSR is of larger amplitude during sound competition. Thus, frequency-selective habituation mechanisms modulate amplitude ratios between 29 and 21 Hz SSR. In the case of pitch continuity, the 21-Hz SSR being highly reduced, the 29-Hz response becomes relatively more important, resulting in an increased saliency of the new-coming 29-Hz component. Conversely, in the case of pitch discontinuity, the 21-Hz response being slightly reduced, the 29-Hz component tends to merge into the acoustic mixture and the two components are grouped into one complex stream. Importantly, this result interpretation is valid if the responses are generated in auditory areas with frequency-selective neurons, but not necessarily tonotopically organized.



plexes (see Alain, 2007 for a review). When all components are tuned, one sound is perceived; whereas when one component is mistuned, two sounds are heard. A negative temporal wave, named “object related negativity” (ORN) has been observed around 180 ms in response to mistuned complex. Even if the amplitude of this wave was correlated with the probability of hearing two distinct sounds, we cannot infer whether this response reflects acoustic or perceptual changes. However, this ORN has also been observed when segregation is induced by inter-aural differences or pitch differences between two vowels. Thus, the ORN does not seem to be specifically related to mistuned harmonic and could more generally index the perception of two different sounds (McDonald and Alain, 2005). However a similar component was not observed in the human auditory cortex when sound segregation was induced by onset asynchrony (Bidet-Caulet et al., 2007A).

NEUROPHYSIOLOGICAL MECHANISMS UNDERLYING SEQUENTIAL ORGANIZATION

Neurophysiological mechanisms involved in auditory sequential organization have been mostly investigated with “streaming” protocol (reviewed in Micheyl et al., 2007; Snyder and Alain, 2007). Animal single-unit, multi-unit, local-field potential (LFP) and human scalp EEG recordings have suggested the involvement of frequency selectivity, **forward suppression** and habituation in sequential organization, in a very similar way as for simultaneous organization.

Forward suppression

A powerful reduction of neuron responsiveness elicited by brief stimuli and that can persist for hundreds of milliseconds, the shorter the delay between two successive stimuli, the stronger the reduction.

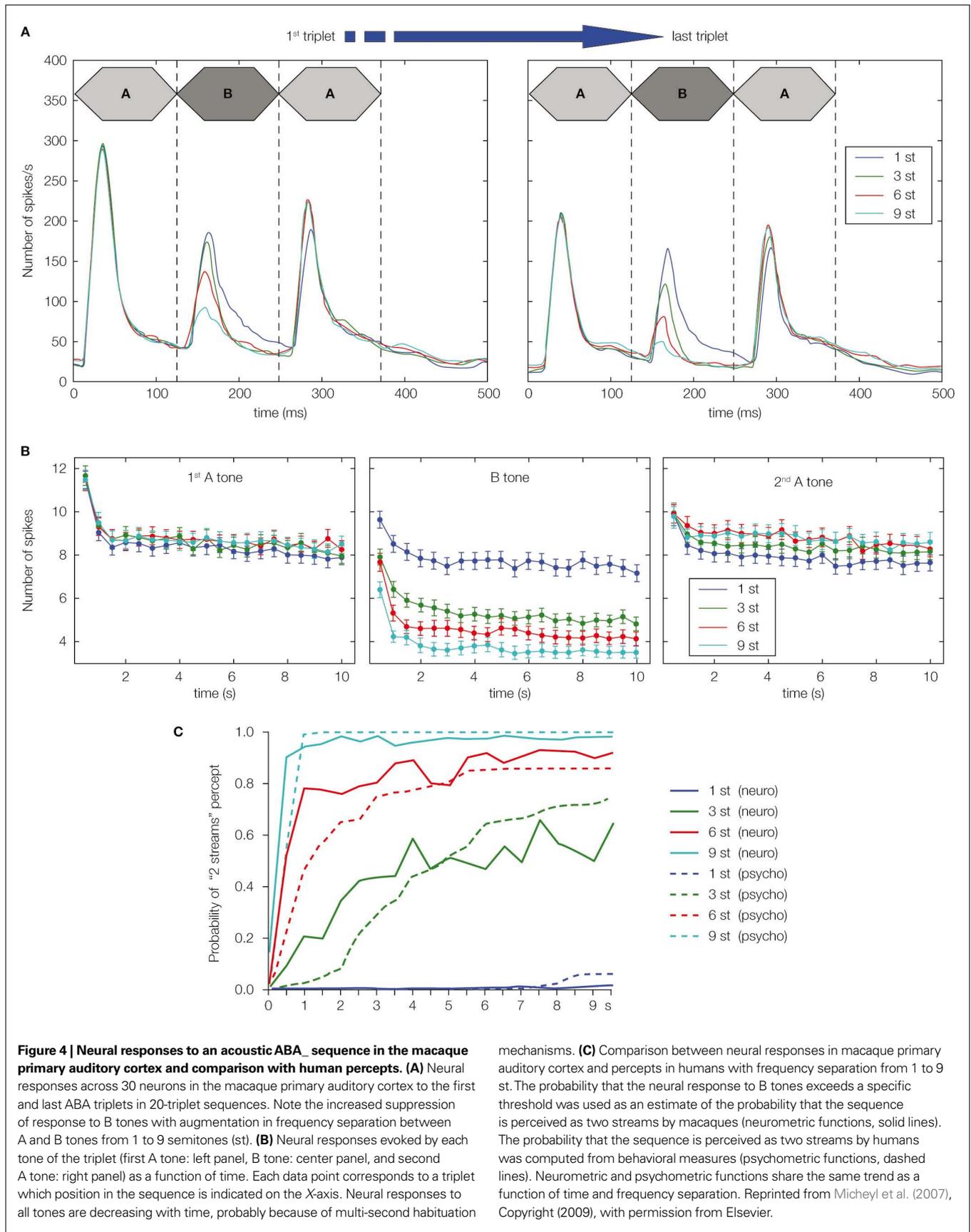
RECORDINGS IN ANIMALS

Frequency-selectivity and forward suppression

Fishman et al. (2001) investigated, for the first time, the neural responses to “streaming” sequences. They recorded both multi-unit activity and LFP from the PAC of awake macaques. The A-tone frequency was adjusted to be close to the best frequency of the recording site, whereas the B-tone frequency differed from A-tone frequency by 10–50%. They found that (1) the faster the tempo, the more the neural response to non-best frequency B tones was attenuated, i.e. the neural activity of the recorded site was mainly composed of responses to best frequency A tones at a rate twice slower than the tone presentation rate, and (2) the more the B-tone frequency was different from the best frequency of the recording site, the stronger the suppression of responses to B tones (see **Figure 4A** for similar results in a more recent study). Thus, the responses found in the PAC were influenced by the tempo and frequency separation of A and B tones in the same way than the percept induced by these sequences in Human.

The authors suggested a forward suppression mechanism which reduces the neural response to a stimulus because of the preceding one, especially when the sounds are close in time. This suppression mechanism is frequency-selective since it is more pronounced for non-best frequency tones than for best frequency tones.

They proposed a physiological model of stream segregation, based on the PAC tonotopic organization:



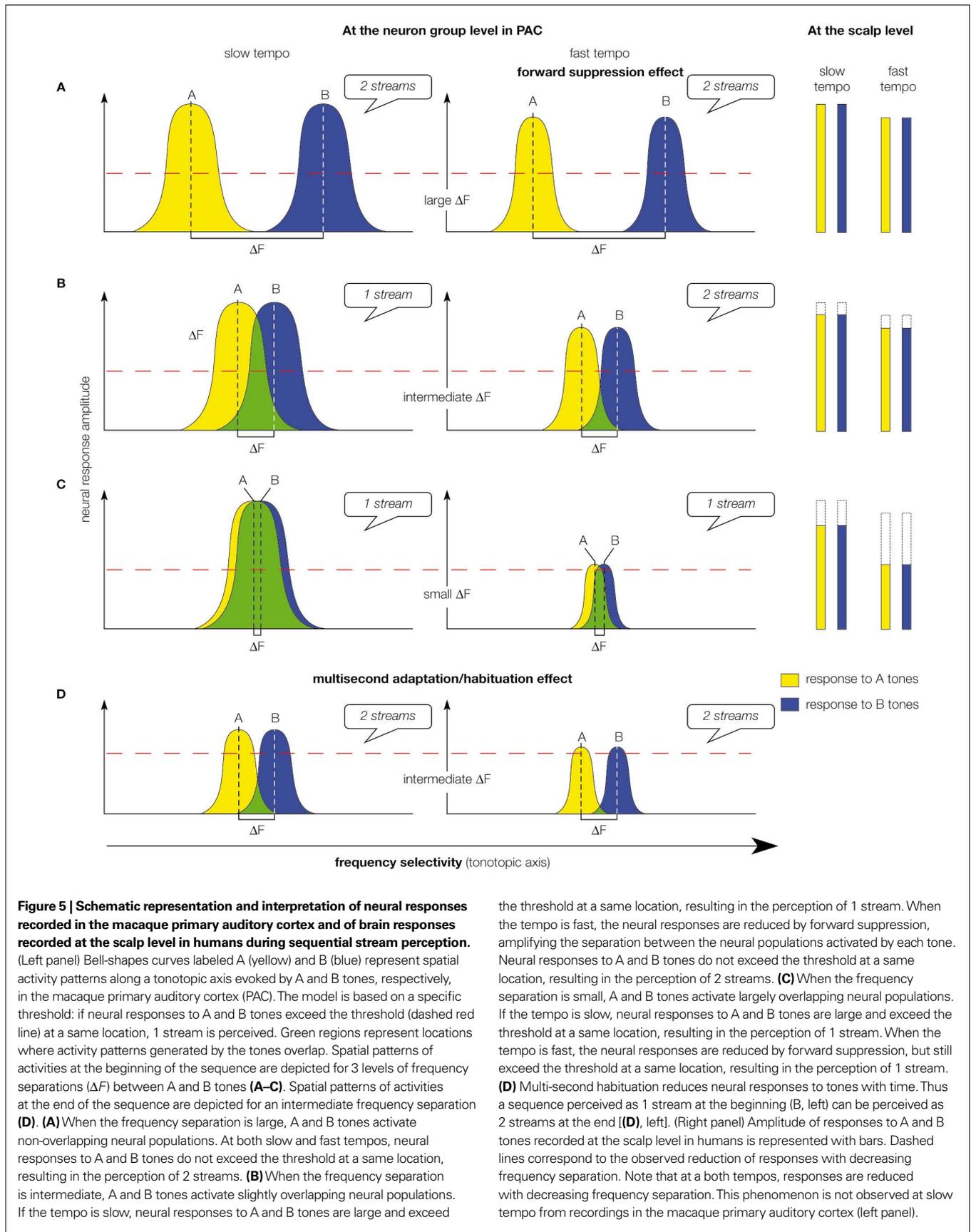


Figure 5 | Schematic representation and interpretation of neural responses recorded in the macaque primary auditory cortex and of brain responses recorded at the scalp level in humans during sequential stream perception. (Left panel) Bell-shaped curves labeled A (yellow) and B (blue) represent spatial activity patterns along a tonotopic axis evoked by A and B tones, respectively, in the macaque primary auditory cortex (PAC). The model is based on a specific threshold: if neural responses to A and B tones exceed the threshold (dashed red line) at a same location, 1 stream is perceived. Green regions represent locations where activity patterns generated by the tones overlap. Spatial patterns of activities at the beginning of the sequence are depicted for 3 levels of frequency separations (ΔF) between A and B tones (**A–C**). Spatial patterns of activities at the end of the sequence are depicted for an intermediate frequency separation (**D**). (**A**) When the frequency separation is large, A and B tones activate non-overlapping neural populations. At both slow and fast tempos, neural responses to A and B tones do not exceed the threshold at a same location, resulting in the perception of 2 streams. (**B**) When the frequency separation is intermediate, A and B tones activate slightly overlapping neural populations. If the tempo is slow, neural responses to A and B tones are large and exceed

the threshold at a same location, resulting in the perception of 1 stream. When the tempo is fast, the neural responses are reduced by forward suppression, amplifying the separation between the neural populations activated by each tone. Neural responses to A and B tones do not exceed the threshold at a same location, resulting in the perception of 2 streams. (**C**) When the frequency separation is small, A and B tones activate largely overlapping neural populations. If the tempo is slow, neural responses to A and B tones are large and exceed the threshold at a same location, resulting in the perception of 1 stream. When the tempo is fast, the neural responses are reduced by forward suppression, but still exceed the threshold at a same location, resulting in the perception of 1 stream. (**D**) Multi-second habituation reduces neural responses to tones with time. Thus a sequence perceived as 1 stream at the beginning (**B**, left) can be perceived as 2 streams at the end [(**D**), left]. (Right panel) Amplitude of responses to A and B tones recorded at the scalp level in humans is represented with bars. Dashed lines correspond to the observed reduction of responses with decreasing frequency separation. Note that at a both tempos, responses are reduced with decreasing frequency separation. This phenomenon is not observed at slow tempo from recordings in the macaque primary auditory cortex (left panel).

- for large frequency separation, A and B tones activate different neural populations, producing the perception of two streams (**Figure 5C**);
- for small frequency separation, A and B tones activate nearly the same neural population, inducing the perception of one stream (**Figure 5A**);
- for intermediate frequency separation (**Figure 5B**), A and B tones activate overlapping neural populations; if the tempo is slow the overlap is large enough to induce a one-stream percept; whereas if it gets faster, suppression mechanism differentially reduces response to best and non-best frequency tones and, consequently, the overlap between neural populations activated by A and B tones, resulting in the perception of two streams.

Therefore, a frequency-selective suppression mechanism amplifies the spatial separation between neural populations activated by A and B tones by narrowing the neuron receptive fields. The more separated these populations, the more likely the sequence would be perceived as two distinct streams.

Fishman's findings have been replicated in the macaque (Fishman et al., 2004; Micheyl et al., 2005), the mustached bat (Kanwal et al., 2003) and the bird (Bee and Klump, 2004, 2005) auditory cortices.

Neural adaptation or habituation

To investigate the buildup of stream segregation (increased probability to hear two streams as the sequence progresses), Micheyl et al. (2005) recorded single unit activity in the PAC of the awake macaque in response to ABA triplets with different frequency separations. They replicated Fishman et al. (2001) results by showing a reduction of firing rate to B tone with increasing frequency separation (**Figure 4A**). More interestingly, they observed that responses to all tones were decreasing from the first to the last triplet embedded in the same sequence, irrespective of the frequency separation. The responses were decreasing strongly during the first 2 s of the sequence and then slowly until the end of the sequence (**Figure 4B**), suggesting the involvement of multi-second adaptation or habituation mechanisms.

Micheyl et al. (2005) computed the probability that the response to B tones (recorded from neurons tuned to A-tone frequency) exceeds a specific threshold and used this value as an estimate of the probability that the same sequence is perceived as two streams in the macaque. They compared this estimate with the probability that the sequence is perceived as two streams

in Human (behavioral measure). They found that both probabilities were showing the same trend as a function of frequency separation and time (**Figure 4C**). This model extends the one by Fishman et al. (2001) by adding habituation mechanisms which reduce responses to A and B tones over time: responses exceeding the threshold at the beginning of the sequence can be reduced below the threshold, leading to a percept switch without any concomitant change in the acoustic content (**Figures 5B,D**).

Interestingly, these results were replicated using recordings from the cochlear nucleus of anaesthetized guinea pig (Pressnitzer et al., 2008): neural responses displayed frequency selectivity, forward suppression and multi-second adaptation and predicted the perception of a sequence. These results were also true for the bushy neurons of the cochlear nucleus receiving direct input from the auditory nerve, raising the possibility that these mechanisms are already present at the level of the auditory nerve.

These studies highly suggest that multi-second adaptation or habituation plays an important role in the switch from one-stream to two-stream percept corresponding to the buildup of streaming. This is consistent with behavioral results in Human showing that the percept is influenced, over several seconds, by previous sounds perceived as one or two streams (Snyder et al., 2008). However, these results do not explain why the perception of the same sequence can spontaneously alternate between one and two streams and *vice-versa* (Pressnitzer and Hupe, 2006). Although perceptual bistability could be explained by bottom-up processes (Hupe et al., 2008; Noest et al., 2007), top-down mechanisms, such as attention, intention or knowledge, might also be involved in these perceptual switches.

Relation between neural and psychoacoustical responses

To build models of stream segregation, different variables were used to predict the percept: the ratio between multi-unit activity to B and A tones (Fishman et al., 2001, 2004), the difference between firing rate to A and B tones (Bee and Klump, 2004) and a threshold (Micheyl et al., 2005; Pressnitzer et al., 2008). To some extent, all these variables depend from each other. They were used to correlate with different psychophysical measures. Indeed, Bee's variable is well correlated with the fission boundary (when subjects cannot avoid hearing one stream even if they try to perceive two streams), but not with the temporal coherence boundary (when subjects can not avoid hearing two streams even if they try to perceive one stream); whereas

Fishman's variable predicts well both boundaries. Finally, Micheyl's threshold is good to estimate the buildup of streaming. It is also noteworthy that all these models share the assumption that neural activation pattern evoked in A1 are "read out" by other neurons which behave as binary classifiers according to a specific variable.

From these previous models, we propose a model of stream segregation (**Figure 5**) based on the separation between the neural populations activated by each tone and on a specific threshold: (1) if responses to A and B tones exceed this threshold within the same neural population, one stream is perceived, whereas (2) if responses to A and B tones exceed this threshold in non-overlapping neural populations, two streams are heard. Our model seems to be based on frequency selectivity and to only account for the segregation of pure tones with distinct frequencies. However, psychophysical studies have demonstrated that perceptual organization does not only depend on frequency parameters, but also on intensity, location, variations over time... Forward suppression and habituation mechanisms could also be selective to other acoustic properties than frequency and might be a general neural mechanism subserving perceptual organization, amplifying the separation between populations activated by the sounds to segregate. Our model takes into account these possibilities since it is based on the overlap and separation between the neural populations activated by each sound, the more different the sounds, the less overlap between the two neural populations, the more likely two streams are perceived. Therefore, our model extends previous models (mainly based on the responses to A and B sounds of the best-frequency neuron) to the population level.

RECORDINGS IN HUMAN

Human studies have the advantage to compare, in the same subjects, brain responses and percepts induced by the sequences. Two studies in MEG (Gutschalk et al., 2005) and EEG (Snyder et al., 2006) found that evoked responses (P1, N1 and P2) to B tones of the ABA triplet were increasing with larger frequency separation and longer interval between tones (**Figure 5**, right panel), in agreement with previous findings showing that N1 amplitude depends on the frequency separation and time interval between two successive sounds (e.g. Picton et al., 1978). Moreover, for a given tempo, the evoked response enhancement was found to be correlated to the percept; the larger P1, N1 and P2 responses, the more likely two streams are perceived. However, as both the percept and evoked response amplitude depend

on the frequency separation, this correlation does not infer if response modulation reflects acoustic or perceptual changes.

At fast tempo, the evoked response reduction with decreasing frequency separation in human could be directly related to the neural response suppression observed in animal PAC (**Figure 5**). Conversely, at slow tempo, no reduction was observed in animals; whereas human brain responses are reduced with decreasing frequency separation (Gutschalk et al., 2005). This suggests that the modulations of these human brain responses do not only reflect differential encoding in the PAC, and that processing in higher level auditory areas could also account for properties of auditory percept construction.

Interestingly, Gutschalk et al. (2005) measures P1 and N1 amplitudes in response to A and B tones embedded in a same sequence inducing the percept of one or two streams. The N1 and P1 amplitudes in response to B and second A tones were found larger when the subjects reported the percept of two streams rather than one. This result is in the same direction as the finding with acoustic manipulations (frequency separation) and suggests that the relationship between the evoked response amplitudes and the percept are not only driven by physical stimulus changes. However, an influence of attention mechanisms on these responses can not be ruled out. Along this line, Snyder et al. (2006) could find, only in an attentive condition, a component of temporal origin paralleling the buildup of stream segregation over time, observed psychologically.

These enhanced transient responses to two-stream percept for sequential organization could appear in contradiction with the stronger transient responses to one-stream percept for simultaneous organization reported above. However, one should keep in mind that, in the case of concurrent sounds (see **Figure 2**), we are comparing responses to one sound (onset asynchrony) with responses to two sounds (onset synchrony), whereas in the case of alternating sounds we are comparing responses to each sound, separately.

CONCLUSION

It is noteworthy that similar mechanisms, namely frequency selectivity, forward suppression and multi-second habituation, have been found involved in both simultaneous and sequential organizations, with multi-second habituation most likely participating in the maintenance and evolution of the percept over time for both organizations. These basic mechanisms, known to be involved in the processing of acoustic properties, have been observed in the auditory cortex,

and at a stage of auditory processing as early as the cochlear nucleus. These findings only demonstrate that neural responses recorded from the auditory cortex or the cochlear nucleus can account for several important characteristics of auditory organization, but they do not prove that the conscious percepts are actually determined at one of this level. It is more likely that percepts are actually formed at a higher level.

A possible scenario would be that basic mechanisms, such as frequency selectivity, suppression, habituation and others, would shape the automatic responses to acoustic properties in subcortical and cortical auditory areas. Then, the resulting patterns of neural activation would be interpreted in higher level areas and/or at a later processing stage to construct the percept. The most challenging question remains: How are these activity patterns transformed, integrated, read out to yield different percepts? Oscillatory activities constitute a good candidate to achieve this binding and underlie the conscious percept. Some cortical evoked brain responses, most likely to be generated in second-

ary auditory areas, were found to be related to the percept rather than the acoustic content, and could also index the percept (Alain, 2007; Gutschalk et al., 2005; Snyder et al., 2006). As multiple mechanisms seem to interact to shape neural responses, several mechanisms could also be involved in reading and interpreting the resulting patterns.

Taken together these results suggest that various mechanisms at different levels of the auditory pathway are involved in auditory perceptual organization. Further research is needed to elucidate which mechanisms actually underlie conscious percepts. In particular, further investigations of oscillatory activities could bring new insights in auditory perceptual organization mechanisms.

ACKNOWLEDGMENTS

The authors wish to acknowledge the contributions of Catherine Fischer, Françoise Lecaigard and Pierre-Emmanuel Aguera for data recording and analysis on the research presented in the focused review.

REFERENCES

- Alain, C. (2007). Breaking the wave: effects of attention and learning on concurrent sound perception. *Hear Res.* 229, 225–236.
- Bee, M. A., and Klump, G. M. (2004). Primitive auditory stream segregation: a neurophysiological study in the songbird forebrain. *J Neurophysiol.* 92, 1088–1104.
- Bee, M. A., and Klump, G. M. (2005). Auditory stream segregation in the songbird forebrain: effects of time intervals on responses to interleaved tone sequences. *Brain Behav. Evol.* 66, 197–214.
- Bidet-Caulet, A., Fischer, C., Bauchet, F., Aguera, P. E., and Bertrand, O. (2007A). Neural substrate of concurrent sound perception: direct electrophysiological recordings from human auditory cortex. *Front. Hum. Neurosci.* 1, 5.
- Bidet-Caulet, A., Fischer, C., Besle, J., Aguera, P. E., Giard, M. H., and Bertrand, O. (2007B). Effects of selective attention on the electrophysiological representation of concurrent sounds in the human auditory cortex. *J. Neurosci.* 27, 9252–9261.
- Bregman, A. (1990). Auditory Scene Analysis: The Perceptual Organization of Sound. Cambridge, MA, MIT.
- Carlyon, R. P. (2004). How the brain separates sounds. *Trends Cogn. Sci.* 8, 465–471.
- Fishman, Y. I., Arezzo, J. C., and Steinschneider, M. (2004). Auditory stream segregation in monkey auditory cortex: effects of frequency separation, presentation rate, and tone duration. *J. Acoust. Soc. Am.* 116, 1656–1670.
- Fishman, Y. I., Reser, D. H., Arezzo, J. C., and Steinschneider, M. (2001). Neural correlates of auditory stream segregation in primary auditory cortex of the awake monkey. *Hear Res.* 151, 167–187.
- Gutschalk, A., Micheyl, C., Melcher, J. R., Rupp, A., Scherg, M., and Oxenham, A. J. (2005). Neuromagnetic correlates of streaming in human auditory cortex. *J. Neurosci.* 25, 5382–5388.
- Hupe, J. M., Joffo, L. M., and Pressnitzer, D. (2008). Bistability for audiovisual stimuli: Perceptual decision is modality specific. *J. Vis.* 8, 1, 1–15.
- Kanwal, J. S., Medvedev, A. V., and Micheyl, C. (2003). Neurodynamics for auditory stream segregation: tracking sounds in the mustached bat's natural environment. *Netw. Comput. Neural Syst.* 14, 413.
- McDonald, K. L., and Alain, C. (2005). Contribution of harmonicity and location to auditory object formation in free field: evidence from event-related brain potentials. *J. Acoust. Soc. Am.* 118, 1593–1604.
- Micheyl, C., Carlyon, R. P., Gutschalk, A., Melcher, J. R., Oxenham, A. J., Rauschecker, J. P., Tian, B., and Courtenay Wilson, E. (2007). The role of auditory cortex in the formation of auditory streams. *Hear. Res.* 229, 116–131.
- Micheyl, C., Tian, B., Carlyon, R. P., and Rauschecker, J. P. (2005). Perceptual organization of tone sequences in the auditory cortex of awake macaques. *Neuron* 48, 139–148.
- Moore, B. C. J., and Gockel, H. (2002). Factors influencing sequential stream segregation. *Acta Acust. United Acust.* 88, 320–333.
- Noest, A. J., van Ee, R., Nijs, M. M., and van Wezel, R. J. (2007). Percept-choice sequences driven by interrupted ambiguous stimuli: a low-level neural model. *J. Vis.* 7, 10.
- Picton, T. W., Woods, D. L., and Proulx, G. B. (1978). Human auditory sustained potentials. II. Stimulus relationships. *Electroencephalogr. Clin. Neurophysiol.* 45, 198–210.
- Pressnitzer, D., and Hupe, J. M. (2006). Temporal dynamics of auditory and visual bistability reveal common principles of perceptual organization. *Curr. Biol.* 16, 1351–1357.
- Pressnitzer, D., Sayles, M., Micheyl, C., and Winter, I. M. (2008). Perceptual organization of sound begins in the auditory periphery. *Curr. Biol.* 18, 1124–1128.
- Snyder, J. S., and Alain, C. (2007). Toward a neurophysiological theory of auditory stream segregation. *Psychol. Bull.* 133, 780–799.
- Snyder, J. S., Alain, C., and Picton, T. W. (2006). Effects of attention on neuroelectric correlates of auditory stream segregation. *J. Cogn. Neurosci.* 18, 1–13.
- Snyder, J. S., Carter, O. L., Lee, S. K., Hannon, E. E., and Alain, C. (2008). Effects of context on auditory stream segregation. *J. Exp. Psychol. Hum. Percept. Perform.* 34, 1007–1016.
- Tallon-Baudry, C., and Bertrand, O. (1999). Oscillatory gamma activity in humans and its role in object representation. *Trends. Cogn. Sci.* 3, 151–162.
- Vidal, J. R., Chaumon, M., O'Regan, J. K., and Tallon-Baudry, C. (2006). Visual grouping and the focusing of attention induce gamma-band oscillations at different frequencies in human magnetoencephalogram signals. *J. Cogn. Neurosci.* 18, 1850–1862.

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

Received: 22 April 2009; paper pending published: 30 May 2009; accepted: 08 July 2009; published: 15 September 2009.

Citation: *Front. Neurosci.* (2009) 3, 2: 182–191. doi: 10.3389/neuro.01.025.2009
Copyright © 2009 Bidet-Caulet and Bertrand. This is an open-access article subject to an exclusive license agreement between the authors and the Frontiers Research Foundation, which permits unrestricted use, distribution, and reproduction in any medium, provided the original authors and source are credited.