Research paper

The role of auditory cortex in the formation of auditory streams

Christophe Micheyl [a,b,*], Robert P. Carlyon [c], Alexander Gutschalk [d], Jennifer R. Melcher [e], Andrew J. Oxenham [a,b], Josef P. Rauschecker [f], Biao Tian [f], E. Courtenay Wilson [g]

[a] *Department of Psychology, University of Minnesota, USA*
[b] *Research Laboratory of Electronics, MIT, Cambridge, USA*
[c] *MRC Cognition and Brain Sciences Unit, Cambridge, United Kingdom*
[d] *Department of Neurology, University of Heidelberg, Germany*
[e] *Eaton-Peabody Laboratory, Massachusetts Eye and Ear Infirmary, Boston, USA*
[f] *Department of Physiology and Biophysics, Georgetown Institute for Cognitive and Computational Sciences,*
Georgetown University Medical Center, Washington DC, USA
[g] *Speech and Hearing Bioscience and Technology Program, Harvard-MIT Division of Health Sciences and Technology, Cambridge, USA*

Abstract

Auditory streaming refers to the perceptual parsing of acoustic sequences into "streams", which makes it possible for a listener to follow the sounds from a given source amidst other sounds. Streaming is currently regarded as an important function of the auditory system in both humans and animals, crucial for survival in environments that typically contain multiple sound sources. This article reviews recent findings concerning the possible neural mechanisms behind this perceptual phenomenon at the level of the auditory cortex. The first part is devoted to intra-cortical recordings, which provide insight into the neural "micromechanisms" of auditory streaming in the primary auditory cortex (A1). In the second part, recent results obtained using functional magnetic resonance imaging (fMRI) and magnetoencephalography (MEG) in humans, which suggest a contribution from cortical areas other than A1, are presented. Overall, the findings concur to demonstrate that many important features of sequential streaming can be explained relatively simply based on neural responses in the auditory cortex.
© 2007 Elsevier B.V. All rights reserved.

Keywords: Auditory cortex; Auditory scene analysis; Single-unit recordings; Magnetoencephalography; Functional magnetic resonance imaging

1. Introduction

An essential aspect of auditory scene analysis relates to the perceptual organization of acoustic sequences into auditory "streams". Typically, an auditory stream is comprised of sounds arising from a given sound source, or group of sound sources, which can then be selectively attended to, and followed as a separate entity amid other sounds. Familiar examples of auditory streams include the sound of a violin in the orchestra, or that of a speaker in a crowd. The process whereby sequentially presented sound elements (e.g., frequency components) are assigned to different streams is usually referred to as "stream segregation"; the converse process, whereby sound elements are bound into a single stream is known as "stream integration".

The essential aspects of auditory streaming can be demonstrated and studied using sequences of pure tones alternating repeatedly between two frequencies, A and B, forming a repeating ABAB… pattern (as illustrated in Fig. 1, panel a), or an ABA_ABA… pattern (Fig. 1, panel b), where the underscore represents a silent gap. Starting with Miller and Heise (1950), numerous psychoacoustical studies have been performed using these types of tone

---

* Corresponding author. Address: Department of Psychology, University of Minnesota, USA. Tel.: +1 612 626 3291; fax: +1 612 626 3359.
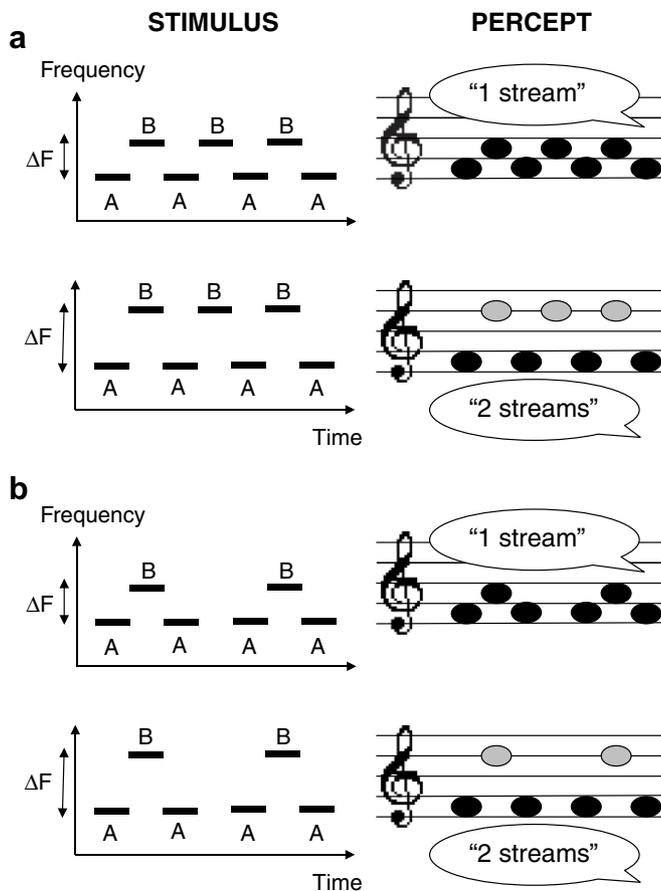*E-mail address:* cmicheyl@umn.edu (C. Micheyl).

Fig. 1. Schematic representation of the stimuli commonly used to study sequential auditory streaming and of the corresponding auditory percepts. The stimulus (left) is a temporal sequence of pure tones alternating between two frequencies, represented here and in the text by the letters A and B. The A–B frequency difference (ΔF) is either small (top left panel) or large (bottom left panel). In the former case, the percept is that of a single, coherent stream of tones alternating in pitch. In the latter case, the percept is that of two separate streams of tones; since the tones in each stream have a constant frequency, the sense of pitch alternation is lost.

sequences (e.g., van Noorden, 1975; Bregman, 1978; Bregman et al., 2000; for a review see: Bregman, 1990; Moore and Gockel, 2002; Carlyon, 2004). One basic result is that, depending on the frequency separation between the A and B tones, the stimulus can evoke two fundamentally different percepts. For relatively small frequency separations, the sequence is usually heard as a single coherent stream of tones alternating in pitch; in the ABA-triplet case, a distinctive "galloping" rhythm is perceived. In contrast, for large frequency separations, and especially if the rate of presentation is fast enough, the sequence is heard as two separate streams: one corresponding to the A tones, the other to the B tones; the sense of pitch alternation (or the galloping rhythm) is lost. Interestingly, over a relatively wide range of intermediate frequency separations, the percept evoked by these stimulus sequences can switch spontaneously from that of a single stream to that of two separate streams, and vice versa, over the course of several seconds or minutes of uninterrupted listening (Bregman, 1978; Anstis and Saida,

1985; Beauvois and Meddis, 1997; Carlyon et al., 2001; Pressnitzer and Hupe, 2006; Kashino et al., in press). However, at these intermediate separations, listeners also appear to have some degree of control over their percept (van Noorden, 1975; but see Pressnitzer and Hupe, 2006).

Using indirect behavioral measures, several investigators have established that the auditory streaming phenomenon can be experienced by various non-human species, including songbird (Hulse et al., 1997; MacDougall-Shackleton et al., 1998), goldfish (Fay, 1998; Fay, 2000), and macaque (Izumi, 2002). This suggests that streaming plays a role in the adaptation to very diverse acoustic environments, or at least that it reflects functional properties of the auditory system that are present in a relatively wide variety of species.

Various theories have been proposed to explain this striking phenomenon of auditory perception. Some of these theories posit the existence of pitch-shift detectors, which are activated by small but not large frequency separations (van Noorden, 1975; Anstis and Saida, 1985); others propose that streaming is determined by proximity in the time–frequency domain (Bregman, 1990), or that it depends on the operation of rhythmic attention (Jones et al., 1981). One theory, which has inspired several computational models (e.g., Beauvois and Meddis, 1996; McCabe and Denham, 1997; Kanwal et al., 2003), posits that stream segregation is experienced when the A and B sounds excite distinct cochlear "filters" or peripheral tonotopic "channels" (van Noorden, 1975); this is commonly known as the "channeling theory". A weaker version of this theory, which was advocated by Hartmann and Johnson (1991), proposes that stream segregation depends on the degree of peripheral tonotopic separation, but does not necessarily require a complete lack of tonotopic overlap. Several psychoacoustical studies have been performed in the past decade, the results of which seriously challenge the channeling theory. In particular, it has been demonstrated that sounds that excite the same cochlear channels can elicit a percept of segregated streams (Vliegen and Oxenham, 1999; Vliegen et al., 1999; Grimault et al., 2000). Moreover, stream segregation can be obtained with sounds that have identical long-term power spectra, based solely on temporal cues (Grimault et al., 2002; Roberts et al., 2002). Conversely, under certain circumstances, sounds presented to different ears, and thus exciting distinct peripheral channels can be grouped into a single stream (e.g., van Noorden, 1975; Deutsch, 1974). Carlyon (2004) has suggested that streaming may take place whenever sounds excite different neural populations, and that these populations can either be located peripherally and be sensitive only to spectral differences, or consist of more central neurons tuned to features such as pitch or temporal envelope. Nonetheless, Hartmann and Johnson's (1991) conclusion that spectral or tonotopic differences provide a strong (and probably the strongest) cue for sequential streaming remains largely undisputed.

While various theories and models have been offered to explain auditory streaming, until recently the actual neural

mechanisms underlying this perceptual phenomenon remained largely unknown and unexplored. However, in the past decade, significant progress has been made in understanding where and how auditory streaming is "implemented" in the brain. This article summarizes the main recent findings in this field. It is divided into two parts. The first is devoted to findings obtained using single- or multi-unit recordings in the primary auditory cortex (A1) of awake primates. The second part presents recent results obtained in studies using magnetoencephalography (MEG), electroencephalography (EEG), or functional magnetic resonance imaging (fMRI) combined with concurrent psychophysical measures in humans. Due to size and scope limitations, some electroencephalographic studies, the findings of which have been interpreted as related to sequential streaming, are not reviewed in the present article. These include in particular studies using the mismatch negativity (MMN) (e.g., Sussman et al., 1999; Alain et al., 1994, 1998) or the T-complex (e.g., Jones et al., 1998; Hung et al., 2001) as indices of streaming. Fortunately, the main findings of these studies have been, or will soon be, summarized elsewhere (e.g., Näätänen et al., 2001; Carlyon and Cusack, 2005; Snyder and Alain, submitted for publication). Instead, the present review focuses on studies in which single-unit recordings or combined physiological and psychophysical measures have been used, in order to gain further insight into the neural mechanisms of streaming in the auditory cortex.

2. Single- and multi-unit recordings: the neural "micromechanisms" of auditory streaming in the primary auditory cortex

2.1. Frequency selectivity and forward suppression in A1 are consistent with the dependence of streaming on frequency separation and inter-tone interval

Although some hints of a possible relationship between neural responses to temporal sound sequences in A1 and sequential streaming can be found in earlier work (e.g., Brosch and Schreiner, 1997), the first published study devoted specifically to this question was performed by Fishman et al. (2001). These authors recorded ensemble neural activity (i.e., multi-unit activity and current-source density) in response to repeating sequences of alternating-frequency tones (ABAB…) in A1 of awake macaques. The frequency of the A tones was always adjusted to be at or near the best frequency (BF) of the current recording site, as estimated from the frequency–response curve. The frequency of the B tones was positioned away from (either below or above) the BF, differing from the A frequency by an amount, ΔF, which ranged from 10% to 50% (i.e., 1.65 to 7 semitones) of the A frequency. Fishman et al. (2001) used short-duration tones (25 ms each, including 5-ms linear ramps), which they presented at rates of 5, 10, 20, and 40 Hz; these tone presentation rates (PRs) correspond to stimulus-onset asynchronies (SOAs, i.e., time between onsets of successive sounds)

of 250, 100, 50, and 25 ms, and to inter-tone intervals (ITIs, i.e., time between the offset of one sound and the onset of the next) of 225, 75, 25, and 0 ms, respectively.

Fishman et al.'s (2001) main findings can be summarized as follows. At slow PRs (5 or 10 Hz), cortical sites with a BF corresponding to A showed marked responses to both the A and the B tones, so that the temporal pattern of neural activity at these sites fluctuated at a rate equal to the PR. In contrast, at fast PRs (20 or 40 Hz), the responses to the B tones were markedly reduced, and the neural activity patterns consisted only or predominantly of the A-tone responses; as a result, the neural activity patterns fluctuated at a rate corresponding to that of the A tones, or half the overall PR.

Fishman et al. (2001) explained these results in terms of physiological "forward masking", where "forward masking" refers to a reduction in the neural response to a stimulus by a preceding stimulus. This phenomenon has been observed in A1 in other studies using two-sound sequences (e.g., Calford and Semple, 1995; Brosch and Schreiner, 1997; Wehr and Zador, 2005). It has been variously described as "monaural inhibition" (Calford and Semple, 1995), "forward inhibition" (Brosch and Schreiner, 1997), or "forward suppression" (Wehr and Zador, 2005). The exact neural mechanisms underlying forward suppression in A1 remain uncertain. While postsynaptic GABAergic inhibition was initially offered as the most likely basis for the phenomenon (Calford and Semple, 1995; Brosch and Schreiner, 1997), it was more recently pointed out that synaptic depression in either thalamocortical or intracortical synapses could also play a role (Eggermont, 1999; Denham, 2001). Consistent with this suggestion, results by Wehr and Zador (2005) implicate intracortically generated synaptic depression, especially at relatively long (>100–150 ms) delays.

Forward suppression in A1 exhibits two important features, as pointed out by Fishman et al. (2001): firstly, the effect increases as the temporal interval (SOA or ITI) between successive tones decreases; secondly, it is more pronounced for non-BF tones than for BF tones (i.e., the responses to non-BF tones is more suppressed by a preceding tone at BF than the other way around). These two features concur to produce a marked suppression of the responses to the B tones by the preceding A tones at fast PRs, while the responses to the A tones are less affected. Fishman et al. (2001) pointed out that the dependence of neural responses in A1 on ΔF and PR, as observed in their study, paralleled the dependence of sequential streaming on these two parameters, as documented in the psychoacoustical literature. Specifically, they noted that at fast PRs, where, according to the psychoacoustical literature, the tone sequences are likely to evoke a percept of two separate streams, neural activity patterns consisted predominantly or exclusively of the neural responses to tones at the BF. Thus, at high PRs, the A and B tones excited distinct neural populations in A1. In contrast, at slow PRs, where the stimulus sequence was presumably perceived as a single stream,

the A and the B tones excited largely overlapping populations, evident as a response to both BF and non-BF tones. These observations are consistent with tonotopic channeling theories and models of streaming wherein perceived segregation depends primarily on tonotopic separation within the auditory system (van Noorden, 1975; Hartmann and Johnson, 1991; Beauvois and Meddis, 1996; McCabe and Denham, 1997). The novel element is that, whereas these theories and models were formulated in reference to the peripheral auditory system, Fishman et al.'s (2001) results relate to the auditory cortex.

Fishman et al.'s (2001) basic findings have been replicated in other species, including the mustached bat (Kanwal et al., 2003) and the European starling (Bee and Klump, 2004, 2005). In particular, Bee and Klump (2004) carried out an extensive study in which they explored systematically the influence of frequency separation, tone duration, and tone PR on neural responses to sequences of tone triplets (ABA_ABA…) in the tonotopically organized field L2 on the auditory forebrain of European starlings – the L2 field may be thought of as the avian equivalent of the mammalian primary auditory cortex. The results of that study are basically consistent with those of Fishman et al. (2001) in showing that increases in frequency separation and/or inter-tone interval promote the forward suppression of the responses to non-BF tones by BF tones. In two more recent studies, Fishman et al. (2004) and Bee and Klump (2005) examined in more detail how neural responses to tone sequences in A1 depended on different temporal parameters of stimulation. In particular, they varied tone duration and tone PR independently in order to assess the respective influence of the tone SOA (time between successive tone onsets) and ITI (time between an offset and the next onset). The results of both studies agree in indicating that ITI, rather than PR, is the determining factor, which is consistent with psychophysical results (Bregman et al., 2000).

2.2. Taking the build-up of streaming into account

The studies by Fishman et al. (2001, 2004) and Bee and Klump (2004, 2005) reveal that neural response properties in A1 are, for the most part, consistent with the dependence of sequential streaming on frequency separation and inter-tone interval. However, this is not the whole story. The percept evoked by a sequence of alternating tones depends not only on these two parameters, but also on the time since the sequence started: at first, the sequence is usually perceived as a single stream, but after a few seconds, it "splits" perceptually into two separate streams; in other words, stream segregation usually takes time (Bregman, 1978; Anstis and Saida, 1985; Beauvois and Meddis, 1997; Carlyon et al., 2001). This tendency for stream segregation to "build up" over time is revealed by having listeners indicate when they start to hear the stimulus sequence as "two streams", and plotting the proportion of "two streams" responses as a function of time since sequence onset, as was done in Fig. 2. The different curves in that figure correspond to different

frequency separations between the A and B tones (1, 6, 3, and 9 semitones). It can be seen that at the smallest frequency separation tested (1 semitone), the listeners almost never perceived the sequence as two separate streams. However, as the frequency separation between the A and B tones increased, the probability of experiencing segregation also increased. At the largest separation tested (9 semitones), the asymptotic probability of a "two stream" response rose very rapidly following the onset of the sequence, and remained high thereafter. At the intermediate separations (3 and 6 semitones), the build-up was slower and/or the asymptotic level lower.

The "build up" of stream segregation can be used to test whether neural responses in A1 (or at any other stage of the auditory system) are consistent with the perceptual experience. If they are, the trend for the percept to change from that of one coherent stream to that of two separate streams should be paralleled by a trend for neural responses to also change in a way that is consistent with the change in percept. Specifically, if the conclusion of the single- and multi-unit recording studies is correct, that perceived segregation depends on the contrast (ratio or difference) between the responses to the A and B tones, under stimulation conditions where the build-up is observed, one would expect an increase in the A–B response contrast over several seconds following the onset of the stimulus sequence. The study
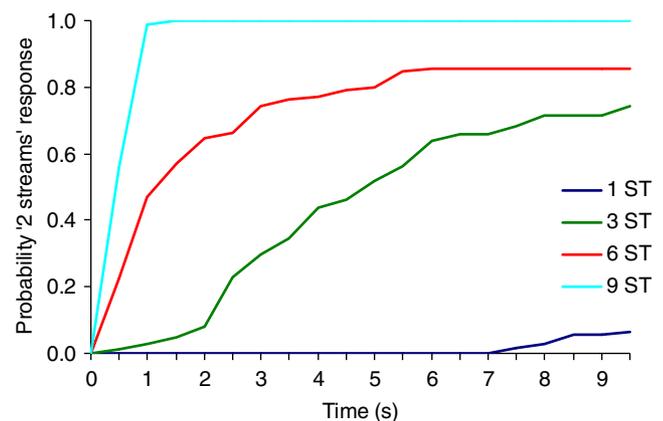

Fig. 2. Example psychometric functions measured in two human subjects and showing how the probability that a repeating sequence of tone triplets (ABA) is perceived as "two streams" varies as a function of time since sequence onset, for different A–B frequency separations. In order to obtain these functions, listeners were presented 20 times with the same 10-s sequence of ABA triplets and they were instructed to report their initial percept (immediately after sequence onset) and any subsequent change in percept (from one to two streams or vice versa), throughout the 10 s. Thus, at a given instant, the response was binary, i.e., one stream or two streams. The probabilities of "two streams" responses were estimated as the measured proportion of trials (out of the 20 trials per condition per listener) on which the listener's percept was that of two separate streams (at the considered instant), averaged across the two listeners. Data corresponding to different frequency separations, from 1 to 9 semitones (ST) are represented by different colors, as indicated in the legend (For interpretation of the references in color in this figure legend, the reader is referred to the web version of this article).

described below (Micheyl et al., 2005), was designed to test this prediction.

Sound sequences similar to those used to measure the psychometric "build up" functions in Fig. 2 were played to two awake macaque monkeys (*Macaca mulatta*), and responses from A1 neurons were recorded. Each sequence consisted of 20 tone triplets (ABA), where each tone was 125 ms long and consecutive triplets were separated by a 125-ms silent gap, leading to an overall sequence duration of 10 s. The only difference in the stimuli between the psychophysical and the electrophysiological experiments was that the frequency of the A tones (relative to which that of the B tones was set) was adjusted to match the estimated best frequency of the unit that was being measured.

Example post-stimulus-time histograms (PSTHs) of neural responses to the first and last ABA triplets in the 10-s stimulus sequences are shown in Fig. 3. The data shown in the upper panel of this figure, and in the following two figures, were obtained using a sub-sample of 30 units, which were selected from a larger sample (*N* = 91) based on the strength of their onset responses to the A and B tones in the 1-semitone frequency-separation condition; the data from the larger sample can be found in Micheyl et al. (2005), and show the same general trends as those illustrated here. In particular, the data show a clear trend for the responses to

the B tone to decrease with increasing frequency separation (compare dark- and light-blue curves in each panel). This effect can be explained in terms of neural frequency selectivity: as frequency separation increased, the B tones moved away from the unit's BF, and they evoked weaker responses. However, additional data shown in Micheyl et al. (2005), who measured responses to the B tones in the absence of the A tones, indicate that forward suppression of the B-tone responses by the preceding A tone also played a role, consistent with the observations of Fishman et al. (2001, 2004). A second trend in Fig. 3 is a decrease in response amplitude between the first and last triplets in the 10-s sequences (compare left and right panels). This second effect indicates a form of response 'adaptation' or 'habituation'. Note that, as illustrated in the lower two panels of Fig. 3, which show PSTHs for a single unit, these effects were already visible in single-unit data in some units.

In order to quantify these effects, we counted the number of spikes in time windows corresponding to the different A and B tones in the different conditions. Fig. 4 shows the average number of spikes produced in response to the A and B tones in each triplet as a function of time. As can be seen, these data confirm the trends observed in Fig. 3: a decrease in the responses to the B tones with increasing frequency separation, and a general decrease in response (for
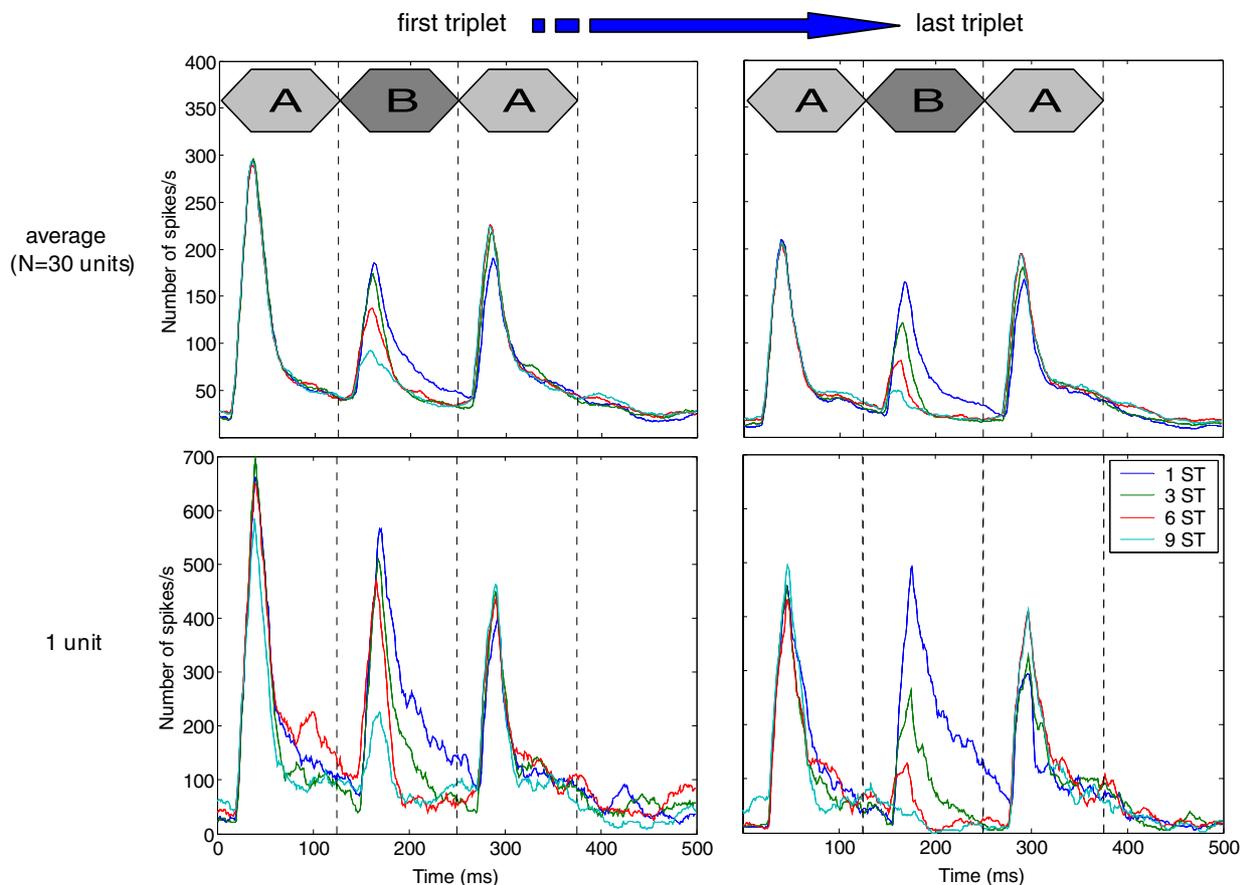

Fig. 3. Post-stimulus time histograms (PSTHs) of neural responses in A1 to the first and last ABA tone-triplets in 20-triplet (10-s overall duration) sequences. The top row shows average PSTHs across 30 A1 units; the bottom row shows the responses from a single A1 unit.

both the A and the B tones) as a function of time since sequence onset. In most cases, the decrease in spike counts was more marked during the first 2 s following sequence onset than during later epochs. However, in some cases, such as for the B-tone at the 3-semitone separation (green curve in the middle panel), the decrease in spike counts extended beyond 2 s.

The decrease in neural responses over time, which is apparent in Figs. 3 and 4, is indicative of neural adaptation in A1. The finding that neural responses to tones adapt in A1 is not unprecedented. Ulanovsky et al. (2003, 2004) investigated the characteristics of spike-count adaptation to tone sequences in cat A1. Their results reveal that adaptation effects at this level of the auditory system range from hundreds of milliseconds to tens of seconds. The results shown in Fig. 4, and those presented by Micheyl et al. (2005), are consistent with those of Ulanovsky et al. in showing spike-count adaptation to tone sequences in A1. Based on Ulanovsky et al.'s finding that neural adaptation in A1 is stronger for more frequent stimuli than for less frequent ones, one might have expected stronger adaptation to the A tones than to the B tones, since with the ABA sequences used here, the A tones were twice as frequent as the B tones. That this expectation was not met in the data is perhaps not surprising, considering that the stimulus conditions tested in the present study differ in several ways from those tested by Ulanovsky et al., and that neural adaptation is likely to be influenced by factors other than just the relative rate of occurrence of the stimuli. In particular, the position of the tones relative to the unit's BF, and the fact that the leading A tones were always preceded by a silent inter-triplet gap whereas the B tones were not, may have influenced the adaptation differently for the A and B tones.

An interesting question relates to the relationship, if any, between the forward suppression phenomenon and the longer-term 'adaptation' or 'habituation' effect. A possible link between the two phenomena is suggested by Wehr and Zador's (2005) whole-cell recording results, which implicate synaptic depression in forward suppression. From this point of view, it is conceivable that the short-term 'suppression' and the longer-term 'adaptation' might both be related to synaptic depression. If this hypothesis is correct, one may expect to find some relationship between the amount of forward suppression and the amount or rate of response habituation across different stimulus conditions. Unfortunately, it is unclear at this point exactly what form this relationship should take. The hypothesis that the rate or amount of habituation should be larger in conditions under which there is more forward suppression is not clearly supported by the data shown in Micheyl et al. (2005). Further study and additional data are required in order to elucidate this question.

2.3. Relating neural and psychophysical responses

The data shown in Fig. 4 indicate that, under stimulus conditions where stream segregation "builds up" (in humans), neural responses in the primary auditory cortex (of macaques) "build down", i.e., they adapt. The question is whether and how the latter effect can account for the former. At first glance, the neural data in Fig. 4, which indicate relatively little influence of frequency separation on the rate and amount of adaptation, seem to be inconsistent with the psychophysical data shown in Fig. 2, which demonstrate large variations in the rate and amount of build-up across frequency separations. However, this apparent discrepancy between the neural data and the psychophysical data can be resolved by considering a less direct, but perhaps more meaningful relationship between the two types of data. Specifically, Micheyl et al. (2005) suggested a simple way of transforming spike counts like those illustrated in Fig. 4 into probabilities of "two streams" responses like those shown in Fig. 2. The solution involves counting the number of trials, out of the total number of times a given stimulus sequence was presented, on which the number of spikes evoked by the B tone remained below a given value or "threshold". Because of the variability inherent in neural responses, the same tone at exactly the same frequency and temporal position within the same sequence does not always elicit the exact same number of spikes on different
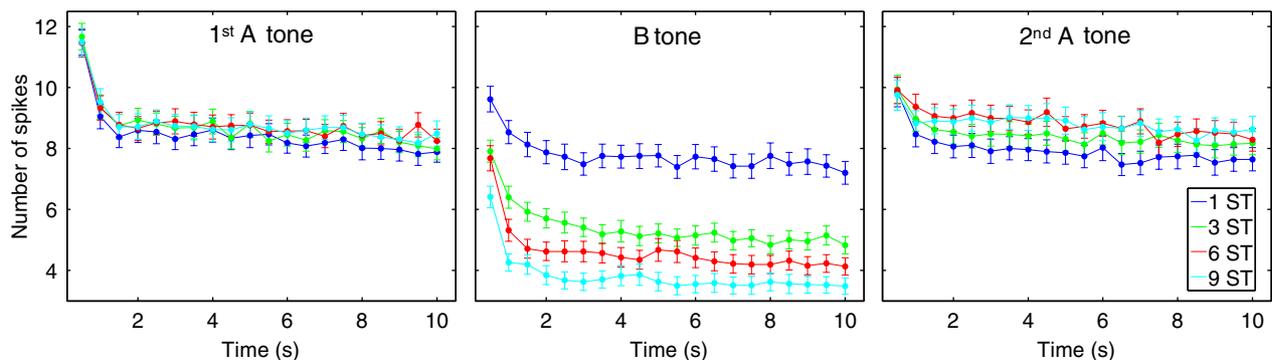

Fig. 4. Number of spikes evoked by the A and B tones in ABA-triplet sequences as a function of time. The three panels correspond to the three tones in each triplet; from left to right: 1st (i.e., leading) A tone, B tone, and 2nd (i.e., trailing) A tone. Each data point corresponds to a triplet. The values along the *X*-axis indicate the onset time of the triplet of which the considered tone was part.

presentations. Therefore, the proportion of trials over which the spike count evoked by the B tone fails to exceed a fixed threshold varies probabilistically. Micheyl et al. (2005) proposed using this probability as an estimate of the proportion of trials leading to a "two streams" response. The idea is that, when the B tone evokes a sub-threshold spike count, only the activation produced by the A tones is detected at the site with a BF corresponding to the A tone. By symmetry, at the site with a BF corresponding to the B tone, only the B tones evoke a supra-threshold response. This response pattern, wherein each tone only excites neurons that have a BF closest to its frequency, corresponds to the case of tonotopic segregation between the A and B tones, which according to the channeling theory of streaming should produce 'two streams' percept. In contrast, cases in which both the A and the B tones evoke supra-threshold counts at both BF sites should elicit a 'one stream' percept.

The scheme outlined above forms the core of the model used by Micheyl et al. (2005) to transform spike trains into perceptual "one stream" or "two streams" judgments. Spike trains are fed into the model, which compares spike counts in consecutive time windows against a pre-determined threshold in order to decide between two possible interpretations of the incoming sensory information. By repeating this process, and with a sufficiently large number of spike-train recordings in response to a given stimulus sequence, one can estimate or "predict" the proportion of times that each perceptual judgment will occur, on average, when this sequence is presented. Another (more computationally efficient) approach for generating these predictions involves estimating the spike-count probability distributions, and then using these to predict the probability that the count falls above the pre-defined threshold. Doing this for each triplet in the stimulus sequence, and for each frequency-separation condition, one can produce "neurometric" functions, which are directly comparable to the psychometric functions shown in Fig. 2. Such neurometric functions are shown in Fig. 5, superimposed onto the psychometric functions from Fig. 2.

Looking at the top panel (Fig. 5a), it can be seen that the neural-based predictions reproduce most of the major trends in the psychophysical data. In particular, they show the expected increase in the probability of segregation with frequency separation and time since sequence onset. Importantly, they capture the interaction between these two effects, showing a slower build-up at intermediate frequency separations (e.g., 3 semitones) than at large ones (e.g., 9 semitones). The good agreement between the neurometric and psychometric functions was confirmed in Micheyl et al. (2005) using statistical criteria: the predictions usually fell within the 95% confidence intervals around the mean data.

Two points regarding the generation of the predictions shown in Fig. 5a are worth mentioning. The first is that the detection threshold in the model was treated as a free parameter. Its value was adjusted to produce the best possible fit between the neural-based predictions and the psychophysical data. However, this value was not allowed to vary with frequency separation or time post sequence onset. Therefore, the dependence of the predictions on these two variables cannot be ascribed to *ad hoc* changes in the position of the threshold; the variation in the predicted proportion of "two streams" responses as a function of frequency separation and time is due entirely to changes in the neural responses. Moreover, although the threshold was treated as a "free" parameter, in practice, the value of this parameter is constrained because if it is set too low, the model will produce an unrealistically high number of false alarms (i.e., erroneous detections during silent periods) due to spontaneous activity, and if it is set too high, the model will fail to detect even tones at the BF.

A second noteworthy point regarding the neurometric predictions illustrated in Fig. 5a is that these were derived by combining spike-count information from a relatively small pool of A1 units ($N = 30$). Interestingly, these predictions already show most of the trends observed in the psychophysical data. In fact, they are not markedly less accurate than those presented in Micheyl et al. (2005), which were based on a larger pool of units ($N = 91$) from which the 30 units used here were selected (based on the strength and onset-like shape of their responses in the 1-semitone condition). This suggests that information from a relatively small number of A1 neurons may be sufficient to account for at least some aspects of perceptual stream segregation. This idea is reminiscent of earlier demonstrations that spike-count information from a few neurons is sufficient to predict perceptual decisions in basic *visual* discrimination tasks (see: Parker and Newsome, 1998). On the other hand, the data shown in Fig. 5b, which show the best-fitting predictions that could be obtained using the spike counts from a single A1 unit, suggest that the psychophysical data cannot quite be accounted for based on truly "single-unit" data.

Finally, it is important to point out that the predictions shown in Fig. 5a and b were obtained under the assumption that the decision variable was based on the spike counts in response to each tone. This is different from what was proposed by Fishman et al. (2001, 2004) and Bee and Klump (2004, 2005), who argued that stream segregation depends on the *contrast* between the neural responses to the A and B tones at a given BF. Fishman et al. (2001, 2004) quantified this contrast by dividing the peak-amplitudes of multi-unit activity or current-source density traces evoked by the A tones to those evoked by the B tones; Bee and Klump (2004, 2005) used the difference between the discharge rates evoked by the A and B tones from the same triplet. Fig. 5c shows best-fitting neurometric predictions obtained using the data collected by Micheyl et al. (2005), under the assumption that stream segregation depends on the difference between the spike counts evoked by the A and B tones in each triplet. In this model, a "two stream" decision was made whenever the spike-count difference exceeded the threshold. These predictions, which were obtained by combining information across all 30 units, do not reproduce the trends in the
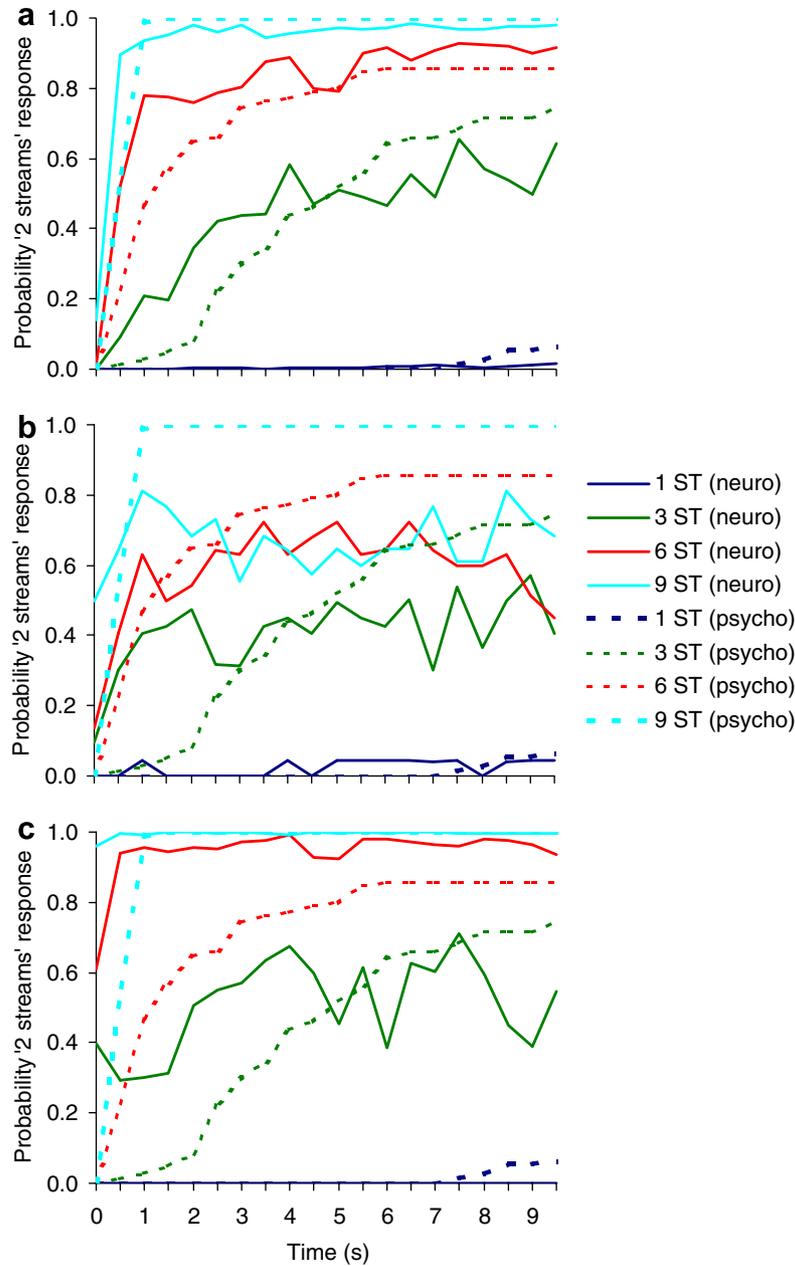
Fig. 5. Comparison between psychometric streaming functions measured in humans and neurometric functions computed based on neural responses measured in monkey A1. The neurometric predictions are shown as solid lines. The psychometric functions from Fig. 2, are re-plotted here using dashed lines in order to facilitate comparison with their neurometric equivalent. (a) Best-fitting predictions derived by combining spike-count information from 30 units. (b) Best-fitting predictions obtained by using spike-count information from a single A1 unit. (c) Best-fitting predictions obtained by combining spike-count information from 30 units, but using as decision variable the *difference* between the spikes counts evoked by the A and B tones in each triplet.

psychophysical data as well as those shown in Fig. 5a; in particular, they over-predict the probability of hearing segregation at the onset of the tone sequence.

2.4. Summary and limitations of the single- and multi-unit studies

Based on the single- and multi-unit studies described above, it is tempting to conclude that several important features of the auditory streaming phenomenon, including its

dependence on frequency separation and inter-tone interval, and the tendency for segregation to build-up over the course of several seconds, can be explained both qualitatively and quantitatively based on neural responses to tone sequences in A1. This conclusion comes with several qualifications. In particular, it should be stressed that these findings only demonstrate that neural responses in A1 can account *in principle* for several important characteristics of streaming; they do not prove that the percepts of streaming are *actually* determined in A1. In fact, the neural-responses

properties that the above studies indicate to be important for streaming, including multi-second adaptation, may already be present below the level of the auditory cortex. Therefore, measuring neural responses to tone sequences similar to those used in psychophysical streaming experiments is an important goal for future studies. Conversely, cortical areas beyond A1 (Rauschecker et al., 1995; Rauschecker and Tian, 2004; Tian and Rauschecker, 2004) may also play an important part in the perceptual organization of tone sequences. Some of the EEG/MEG and fMRI studies described in the following section of this review address this question.

A second limitation of the studies is that, at best, they compared neural responses obtained in one species with psychophysical data collected in another. Ideally, the comparison should be between neural and psychophysical data collected simultaneously in the same subject. Thus, an important goal for future work is to combine single- or multi-unit recordings with behavioral measures of streaming in the same animal, and if possible, at the same time.

Another significant limitation of the above studies is that they only used pure tones. Psychophysical studies performed during the past ten years have demonstrated that stream segregation can occur with other types of stimuli, including sounds that occupy the same spectral region and excite the same peripheral tonotopic channels (e.g., Vliegen and Oxenham, 1999; Vliegen et al., 1999; Grimault et al., 2000). In fact, it has been shown that stream segregation can occur even with A and B sounds that have identical long-term power spectra, such as amplitude-modulated bursts of white noise (e.g., Grimault et al., 2002) or harmonic complexes differing only by the phase relationship of their harmonics (Roberts et al., 2002). These findings challenge explanations of streaming in terms of peripheral tonotopic separation. They call instead for a broader interpretation of the currently available single- and multi-unit data, in terms of functional separation of neural populations; tonotopy is one important determinant of this functional separation, but not the only one. For instance, it has been shown that some neurons in the auditory cortex display non-monontonic rate-level functions, indicating level-selectivity, in addition to frequency-selectivity (e.g., Phillips and Irvine, 1981); such neurons could mediate streaming between tones of the same frequency, based on level differences (van Noorden, 1975). Another example is provided by Bartlett and Wang's (2005) finding that neural responses to amplitude- or frequency-modulated tones or noises in the auditory cortex of awake marmosets can be influenced by a preceding sound, and that these temporal interactions are tuned in the modulation domain. These observations open the possibility that neural response properties in A1 may also account for streaming based on temporal cues, such as differences in amplitude-modulation rate between broadband noises (Grimault et al., 2002) or difference in fundamental-frequency between complex tones containing only peripherally unresolved harmonics (Vliegen and Oxenham, 1999; Vliegen et al., 1999; Grimault et al., 2000).

Finally, several important questions remain, regarding whether and how attention and intentions, which have been shown to influence streaming, also influence neural responses in and below the auditory cortex. For instance, the findings of Carlyon et al. (2001) and Cusack et al. (2004), which indicate that attention can influence the build-up of stream segregation, raise the question as to whether this effect is mediated in, beyond, or below A1. If the effect of attention occurs at the level of A1 or below it, then one should see an influence on multi-second adaptation in A1. A related question regards the influence of the listener's intentions. It has been shown that listeners can bias their percepts toward integration or segregation, depending on what they are trying to hear. Micheyl et al.'s (2005) model, which includes an internal 'threshold' or 'criterion', provides a simple way to capture such effects via changes in the position of the criterion depending on the listener's intention – consistent with the framework of classical psychophysical signal detection theory (Green and Swets, 1966). However, an alternative possibility is that intentions directly affect neural responses in A1 or below. Lastly, if the percept of streaming is determined in or below A1, one should be able to see fluctuations in neural responses inside A1, which correlate with the seemingly spontaneous switches in percept that occur upon prolonged listening to repeating sequences of alternating tones at intermediate frequency separations (Pressnitzer and Hupe, 2006). The inherently stochastic nature of neural responses in the peripheral and central auditory system provides a likely origin for the stochastic fluctuations in percept over time. Unfortunately, not enough information is currently available regarding how neural responses to sound sequences fluctuate over several tens of seconds or more.

3. Magnetoencephalographic (MEG) and functional magnetic resonance imaging (fMRI) measures of auditory streaming in humans

3.1. MEG and EEG correlates of sequential streaming in the human auditory cortex

Gutschalk et al. (2005) and Snyder et al. (2006) recorded magnetic and electric auditory evoked responses, respectively, to sequences of ABA triplets, which they also used in psychophysical measurements of auditory streaming in the same listeners. One basic finding of these two studies was that the magnitude of the response time-locked to the B tones increased with the A–B frequency separation. This effect is illustrated in Fig. 6, which shows average source waveforms evoked by the ABA triplets from an MEG study similar to that described in Gutschalk et al. (2005). As can be seen, each tone in the stimulus (ABA or A_A) gave rise to a transient response characterized by a positive peak with a latency of approximately 50–70 ms relative to the onset of the corresponding tone, followed by a negative peak with a latency of approximately 100–120 ms (relative to tone onset). These two peaks correspond to the well-
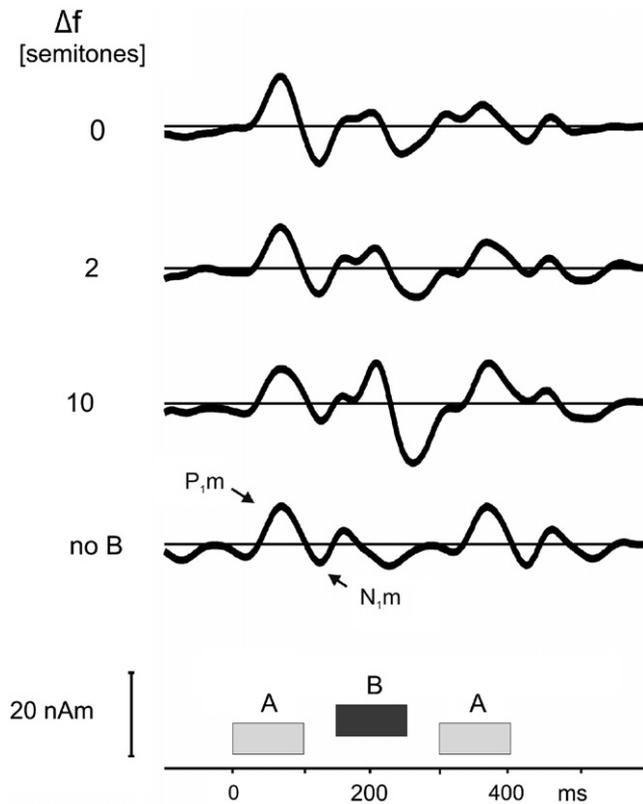
Fig. 6. Average source waveforms from the auditory cortex of 14 listeners in response to ABA tone triplets with different A–B frequency separations (indicated in semitones on the left; they increase from top to bottom), and for a control condition in which the B tone was replaced by a silent gap (bottom trace). The P_1m and N_1m peaks evoked by the trailing A tone are labeled on the bottom trace. The three tones are represented schematically at their respective temporal positions underneath the traces. Each tone was 100 ms (including 10 ms ramps), the offset-to-onset interval-tone interval within each triplet, 50 ms, and the inter-triplet interval, 200 ms. The source waveforms shown here reflect activity originating from dipoles located in or near Heschl's gyrus (right and left hemispheres), as determined by spatiotemporal dipole source analysis (Scherg, 1990). Further methodological details can be found in Gutschalk et al. (2005). The main difference between the experiments described in that earlier study and that reported here is that, here, the A frequency was fixed and the B frequency variable – whereas the converse was true in Gutschalk et al. (2005).

documented P_1m and N_1m peaks of the long-latency magnetic auditory evoked field. Results from a control condition, where the B tones were replaced by silent gaps, are also shown. While for the data presented in Gutschalk et al. (2005), the B-tone frequency was constant, to analyze the effect of ΔF on the B-tone evoked response unrelated to changes of B-tone frequency, the data in Fig. 6 were recorded with constant A-tone frequency, but otherwise similar parameters. Note that the same approach of keeping the A-tone frequency constant was used in the study by Snyder et al. (2006). Overall, the comparison shows that the enhancement of the B-tone response is very similar for both the fixed A- and B-tone conditions; additionally the data in Fig. 6 confirms that some, albeit smaller variations are observed for the A-tones as well: while at ΔF of 0- and 2-semitones, the response evoked by the trailing A tone was

reduced compared to that evoked by the leading A tone, the responses to the two A tones are of roughly equal magnitude at 10-semitones ΔF, similar to the control (no B tone) condition.

These observations, which complement those presented in Gutschalk et al. (2005) and Snyder et al. (2006), are consistent with earlier findings indicating that the magnitude of the N_1 (or N_1m) depends on the frequency separation and the inter-tone interval between consecutive tones (Butler, 1968; Picton et al., 1978; Näätänen et al., 1988). These various findings can be explained in terms of forward-suppression or adaptation, whereby the neural response to a tone is diminished by a preceding tone, when the frequency separation between the tones is small and the inter-tone interval is short, This raises two questions: First, do the frequency-separation-dependent forward suppression or adaptation effects apparent in these EEG and MEG studies share a common basis with those observed at the single-unit level in the single- and multi-unit recording studies reviewed above? Second, are these effects entirely stimulus-driven, or are they related to how the sequence is perceptually organized (i.e., into one or two streams) by the listener?

Regarding the first question, it is worth noting that whereas the single- and multi-unit data were obtained specifically in A1, the MEG and EEG responses that Gutschalk et al. (2005) and Snyder et al. (2006) found to vary with frequency separation most probably reflect activity from multiple, mostly non-primary areas of the human auditory cortex. Intracranial recordings (Liegeois-Chauvel et al., 1994) and source-analysis studies (Pantev et al., 1995; Yvert et al., 2001; Gutschalk et al., 2004) in humans indicate that the P_1 and N_1 originate in the activity of neurons located in the lateral part of Heschl's gyrus, the planum temporale (PT), and the superior temporal gyrus (STG). On the other hand, an analysis of the middle-latency data from Gutschalk et al.'s study (not described in their 2005 article) did not show a significant effect of frequency separation on the P_am and N_am middle-latency peaks, the main generator of which is believed to be in medial Heschl's gyrus (Liegeois-Chauvel et al., 1991), and thus presumably in human A1 (Flechsig, 1908; Galaburda and Sanides, 1980; Hackett et al., 2001; Morosan et al., 2001). A similar re-analysis of Snyder et al.'s (2006) data, which involved a larger number of repetitions (approximately 2000) per condition also failed to show a significant influence of frequency separation (Snyder, personal communication). Although these negative results deserve further scrutiny, they suggest that Gutschalk et al.'s and Snyder et al.'s MEG and EEG results reflect activity originating in different parts of the auditory cortex than explored in the single-unit studies of Fishman et al. (2001, 2004) and Micheyl et al. (2005).

Another obstacle in relating the single and multi-unit data with those obtained in EEG or MEG studies comes from the fact that the unit data reflect local neural responses, which essentially correspond to a single BF site, whereas the EEG or MEG data reflect more global responses that encompass several BF sites. As an example

of this, the reduction in response to the B tones with increasing frequency separation in the single-unit data does not translate into a similar effect in the EEG or MEG data, because as frequency separation increases, different populations of neurons are recruited. Thus, the exact relationship between the currently available single-unit and EEG/MEG data on the neural basis of streaming remains unclear.

Regarding the question of whether the changes in EEG/MEG responses reflect physical or perceptual changes, both studies (Gutschalk et al., 2005; Snyder et al., 2006) provide some indication that the amplitude changes in the long-latency response may in fact reflect changes in perception. For instance, Gutschalk et al. (2005) and Snyder et al. (2006) both found that the magnitudes of the P_1/N_1 or P_{1m}/N_{1m} peaks evoked by the B tones at different A–B frequency separations were statistically correlated with psychophysical measures of streaming in the same listeners. Admittedly, the conclusions that can be drawn on the basis of this result are limited because the co-variation between these two variables could be due merely to their co-dependence on a common underlying factor: frequency separation. Thus, in a second experiment, Gutschalk et al. (2005) measured MEG responses and perceptual judgments concurrently in the same listeners while holding the physical stimulus constant. The subjects listened to sequences with a frequency separation of 4 or 6 semitones, which was chosen to fall into the ambiguous region, where the stimulus could be heard sometimes as a single stream, sometimes as two separate streams. By having listeners monitor their percept throughout the sequence and report each time a switch occurred, Gutschalk et al. (2005) could index the MEG recordings onto the listener's percept, and compare brain responses collected while the percept was "one stream" with responses collected while the percept was "two streams". In this way, the authors were able to look for changes in the brain responses associated with changes in percept, without the confounding factor of stimulus differences. The results showed that the magnitude of the P_1m/N_1m in the response to the B tones was consistently larger when the percept was "two streams" than when it was "one stream". The difference was smaller than, but in the same direction as, that observed in their first experiment, where larger P_1m/N_1m peaks were found in response to the B tones under conditions producing a "two streams" percept. This findings of the second experiment thus suggest that the variation in the magnitude of the MEG responses to the B tones as a function of the A–B frequency separation in the first experiment was not driven solely by *physical* stimulus changes (i.e., changes in frequency separation); *perceived* changes (i.e., changes in pitch separation) also had an influence.

Another argument for the notion that neural responses in auditory cortex are related to the percept of streaming, and not just to the physical parameters of the inducing stimulus sequence, was provided by Snyder et al. (2006). These authors analyzed the EEG responses evoked at different time points after the onset of relatively short (10 s) stimulus sequences. They found that a positive component peaking about 200 ms after the beginning of each ABA triplet, increased over several seconds after sequence onset, paralleling the build-up of stream segregation observed psychophysically.

3.2. fMRI indices of sequential streaming in the human auditory cortex and beyond

In contrast to the MEG and EEG findings, described above, which demonstrate co-variations between the percepts induced by unchanging tone sequences and the neural responses evoked in the auditory cortex by those same sequences in the same listeners, fMRI results by Cusack (2005) show no percept-correlated changes in auditory cortex activation while listening to perceptually ambiguous tone sequences. On the other hand, the results of that study showed that activation within the intraparietal sulcus depended on the listener's percept; the activation in this extra-auditory cortical region was significantly larger when averaged over stimulation epochs during which the listener had reported hearing two streams, compared to that for epochs during which the listener reported hearing a single stream. As a possible explanation for this surprising finding, Cusack (2005) pointed out that the intraparietal sulcus has been implicated in feature binding in the visual domain, and cross-modal integration. In any case, Cusack's (2005) findings indicate that brain areas outside of those traditionally regarded as part of the auditory cortex may be involved in auditory streaming. Whether these areas are actively engaged in the formation of auditory streams or activated following the perceptual organization of the sound sequence into streams, remains to be determined, and offers an interesting avenue for future research.

One surprising aspect of Cusack's (2005) results is that they show no significant effect of frequency separation on activation in the auditory cortex. This was the case even though Cusack (2005) used a relatively wide range of frequency separations (1–7 semitones), over which clear variations in auditory cortex responses across frequency separations were observed in the above MEG and EEG studies. This apparent discrepancy in outcomes might stem from the fact that fMRI and MEG/EEG capture different aspects of neural activity. However, other recent fMRI findings by Wilson et al. (2005) have demonstrated marked variations in auditory cortex activation as a function of the frequency separation between the A and B tones in repeating ABAB sequences. Specifically, Wilson et al. (2005) found that both the amplitude and the spatial extent of activation increased with the A–B frequency separation in two regions of interest (ROIs), one corresponding to the postero-medial two-thirds of Heschl's gyrus (HG), the other to the planum temporale (PT). This result is illustrated in Fig. 7a, which shows increasing activation on the superior temporal lobe of one listener. The marked increase in auditory cortex activation with increasing frequency separation between the A and B tones, which is apparent in these individual data, is typical of that observed by

Wilson et al. (2005). The reason for the discrepancy in outcomes between Cusack's (2005) and Wilson et al.'s (2005) studies regarding the influence of A–B frequency separation on responses in auditory cortex remains unclear at this point. It could be due to differences in stimuli. In particular, whereas Cusack used sequences of ABA triplets separated by silent gaps, Wilson et al. used ABAB sequences with no silent gaps, which were selected to induce markedly different temporal patterns of activation in auditory cortex depending on the A–B frequency separation.

In addition to looking at the overall magnitude and extent of auditory cortex activation, Wilson et al. (2005) also analyzed the time course of activation evoked by the 32-s sequences of repeating A and B tones. Using a generalized linear model, they fitted basis functions to the data (Harms and Melcher, 2003) and quantified the time course of activation using a summary "waveshape index" (Harms and Melcher, 2003) comprised between 0 (indicating highly sustained or "boxcar" activation) and 1 (indicating a completely phasic activation pattern consisting only of peaks following the onset and offset of the sequence, with no sustained activity in-between). They found that as frequency separation increased, the activation generally became more sustained (i.e., less phasic). This is illustrated in Fig. 7b,

which shows the time course of auditory cortex activation for different A–B frequency separations. Finally, the activation produced by ABAB sequences with a frequency separation of 20 semitones was similar in shape to that evoked by control sequences consisting only of the A or only of the B tones: in all these cases, the activation was largely sustained, contrasting with the more phasic activation obtained in the 0-semitone condition.

The interpretation proposed by Wilson et al. (2005) to explain these findings relies on the same basic forward suppression mechanism invoked by Gutschalk et al. (2005) to explain their MEG data, and by Fishman et al. (2001, 2004) and others to explain their single- and multi-unit recordings. The idea is that, when the A and B tones are close in frequency and occur in close temporal succession, the neural response to each tone is suppressed by the preceding tone, the result being a reduction in the magnitude of fMRI activation with decreasing frequency separation, as well as a more rapid decline in activity just after sequence onset (leading to a more phasic response). The latter result strongly resembles changes in activation time course that occur when the physical repetition rate of sounds in a sequence is increased (Harms and Melcher, 2002; Harms et al., 2005). When the rate of sound presentation is slow
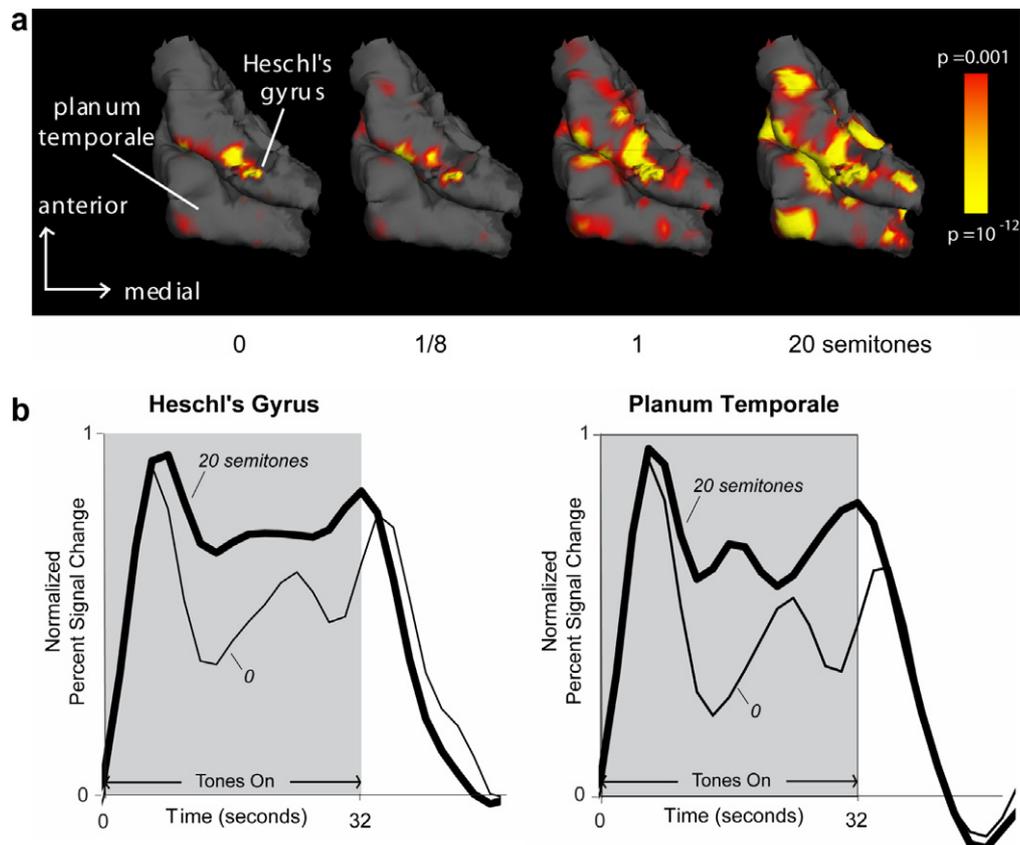

Fig. 7. (a) Auditory cortex activation in response to repeating ABAB sequences for different A–B frequency separations. The activation is shown overlaid on a 3D reconstruction of the superior temporal lobe obtained from T1-weighted (~1 × 1 × 1 mm resolution) images (MRPAGE). Stronger activation, corresponding to lower statistical p values, appears in bright yellow; weaker (albeit statistically significant) activation, in red. The data shown here are from a single listener, but typical of those obtained in a larger sample (Wilson et al., 2005). (b) Time courses of activation in auditory cortex for the tow extreme A–B frequency separations: 0 and 20 semitones (For interpretation of the references in color in this figure legend, the reader is referred to the web version of this article.).

(gap between sounds greater than about 200 ms), the activation of auditory cortex is largely sustained throughout a sequence; however, as the rate is increased (silent gaps shortened), activation becomes more phasic, displaying prominent peaks at stimulus onset and offset. The shift in time course between sustained and phasic has been shown to occur specifically for changes in the temporal properties of sound (e.g., rate), rather than for changes in bandwidth or level (Harms et al., 2005). This specificity led Wilson et al. (2005) to suggest that the observed time course changes with frequency separation may be directly related to the co-occurring changes in the perceived rate of their stimulus sequences: slow for large frequency separations and fast for small separations. One interesting possibility is that auditory cortex responds to the perceived rate of sound events rather than (or in addition to) the physical rate of stimulation. The fact that perceived rate is intimately linked to the number of perceived streams (two for slow rates, one for fast) suggests that the measured dependences of fMRI activation time course on frequency separation may also be directly related to the perceptual organization of the sequences.

While the nature of the time course changes in Wilson et al.'s (2005) study suggests a direct relationship to listeners' perceptions of the sequences, it must be acknowledged that these data are by no means decisive. In fact, they leave open the possibility that the different patterns of auditory cortex activation observed at small and large frequency separations might be completely unrelated to the listener's actual percepts of the stimulus sequence as one or two streams. In order to rule out this possibility, it would be necessary, in a second step, to demonstrate co-variations in the neural and psychophysical responses in the absence of concomitant changes in the physical stimulus, as was done in the MEG study of Gutschalk et al. (2005) or the fMRI study of Cusack (2005).

3.3. MEG and fMRI correlates of non-tonotopically based streaming

As mentioned earlier, it is now well established that stream segregation occurs in situations where consecutive sounds excite the same peripheral channels and evoke similar or identical tonotopic excitation patterns inside the auditory system (e.g., Vliegen and Oxenham, 1999; Vliegen et al., 1999; Grimault et al., 2000, 2002; Roberts et al., 2002), arguing against the tonotopic channeling theory of streaming. If the neural phenomena identified in the above MEG and fMRI studies are truly related to streaming, they may also manifest themselves under stimulation conditions that do not involve spectral differences. In all of the above-described MEG and fMRI studies, the stimuli were sequences of pure tones and stream segregation was induced by frequency differences, which produced salient tonotopic cues. A recent fMRI study by Deike et al. (2004) used ABAB sequences in which the A and B sounds were harmonic complex tones that spanned the same frequency range but differed in spectral envelope – and thus timbre. The results of that study are consistent with those of Wilson et al. (2005) in showing stronger activation in auditory cortex in response to such alternating sequences, which were perceived as two streams, compared to sequences consisting of identical sounds (AAAA or BBBB), which were perceived as a single stream. However, this does not fully address the question of whether differences in auditory cortex activation can be observed in situations where perceptual organization cannot be based on tonotopic cues, because the complex tones used as A and B sounds by Deike et al. (2004) did evoke different tonotopic patterns of excitation.

In order to overcome this limitation, Gutschalk et al. (2006) recently carried out MEG and fMRI experiments using complex tones that differed in fundamental frequency (F0) but had identical spectral envelopes and were band-pass-filtered into a relatively high spectral region, in such a way that their harmonics were not spectrally "resolved" in the cochlea (Carlyon and Shackleton, 1994)); thus, these sounds differed noticeably in pitch but presumably evoked very similar tonotopic patterns of excitation in the auditory system. Interestingly, their results show similar effects to those observed in the earlier MEG and fMRI studies by Gutschalk et al. (2005) and Wilson et al. (2005), which both used pure tones. Thus, Gutschalk et al.'s (2006) results provide further evidence that the effects observed by Gutschalk et al. (2005) and Wilson et al. (2005) were not mediated by mechanisms that only operate along the tonotopic dimension. As mentioned above, recent observations based on single-unit recordings in marmoset A1 reveal the existence of temporal interactions in the neural responses to consecutive sounds, which are at least superficially similar to the temporal interactions previously observed with pure tone sequences (i.e., forward suppression or enhancement), but also manifest themselves with other types of stimuli (such as noises), and depend on stimulus parameters other than carrier frequency (such as amplitude-modulation rate) (Bartlett and Wang, 2005). This suggests that neural mechanisms analogous to those invoked to explain auditory streaming for pure tones, such as selectivity to stimulus parameters and forward suppression or adaptation, operate along other axes of stimulus representation than tonotopy, and might explain auditory streaming based on parameters other than frequency. Interestingly, consistent with earlier findings indicating a specialized pitch area in humans auditory cortex (Griffiths et al., 1998; Gutschalk et al., 2002; Patterson et al., 2002; Krumbholz et al., 2003; Penagos et al., 2004), pitch-selective neurons have recently been identified in an antero-lateral region of the auditory cortex of monkeys (Bendor and Wang, 2005). Such specialized neurons, and other neurons that are specifically sensitive to complex stimulus features, may mediate changes in the patterns of activation in the auditory cortex depending on complex stimulus features, which also contribute to the perceptual organization of sound sequences into streams.

4. Conclusions and perspectives

The single- and multi-unit, MEG, and fMRI studies reviewed above are all consistent with the idea that several important features of the perceptual organization of repeating tone sequences into auditory streams can be explained relatively simply, based on neural responses to these sequences in the auditory cortex. Specifically, the single- and multi-unit data reveal that the characteristics of frequency selectivity, forward suppression, and multi-second adaptation of neural responses measured in the primary auditory cortex, can account qualitatively and quantitatively for psychophysical measures of pure-tone stream segregation as a function of frequency separation, inter-tone interval, and time since sequence onset. The MEG, EEG, and some of the fMRI data currently available, further support an involvement of frequency selectivity, forward suppression, and adaptation in pure-tone streaming. However, they also qualify the single- and multi-unit results by revealing that superficially similar neural phenomena are observed in non-primary areas of the auditory cortex as well. Furthermore, recent MEG and fMRI studies using complex rather than pure tones suggest that neural selectivity to other attributes than frequency, such as fundamental frequency (virtual pitch) or amplitude modulation, combined with forward suppression or selective adaptation between neural populations tuned to such dimensions in the auditory cortex, may mediate stream segregation for sounds that are not well separated along the tonotopic axis. Finally, in addition to showing that neural responses in the auditory cortex are *consistent with* psychophysical measures of streaming, some of the above-described MEG and EEG studies demonstrate *co-variations* between neural and psychophysical responses measured concurrently in the same listeners. These co-variations, especially those seen while holding physical stimulus constant, provide further support to the notion that the perceived organization of sound sequences is reflected in neural activity in the auditory cortex, and that this activity does not just depend on the physical stimulus parameters.

Many important questions remain. In particular, all experimental studies on the neural basis of streaming so far have focused exclusively on the auditory cortex (or its equivalent in non-mammalian species), leaving largely open the question of whether neural responses at lower stages of the auditory system are also compatible, qualitatively and quantitatively, with psychophysical measures of stream segregation. Therefore, an important goal for future studies of the neural basis of auditory streaming is to measure responses to sequences of alternating tones at different stages of auditory processing, in order to determine to what extent neural responses at each stage can or cannot account for the phenomenon. Ideally, such investigations should be coupled with behavioral measures of streaming obtained in the same species and under the exact same stimulus conditions.

Another important goal for future studies is to investigate the influence of attention and of the listener's intentions on neural responses to tone sequences at various stages of the auditory system, in order to determine at what stage, and via which mechanisms, the influence of such factors on streaming is determined. Therefore, neural responses to alternating tone sequences should be systematically measured under different attentional conditions as well as in the context of different tasks, some promoting segregation of the sequence into separate streams, others promoting integration into a single stream.

Lastly, a better understanding of the neural mechanisms of sequential auditory scene analysis requires that research in this area progressively moves away from stereotyped pure-tone sequences, toward spectro-temporally complex and partly random sound sequences more typical of those found in the environment. The fMRI and MEG studies described above, which used harmonic complex tones, are a step in this direction, as are more recent exploratory studies in humans and animals (Micheyl et al., 2007; Yin et al., in press).

Acknowledgements

References

Alain, C., Woods, D.L., Ogawa, K.H., 1994. Brain indices of automatic pattern processing. Neuroreport 30, 140–144.

Alain, C., Cortese, F., Picton, T.W., 1998. Event-related brain activity associated with auditory pattern processing. Neuroreport 15, 3537–3541.

Anstis, S., Saida, S., 1985. Adaptation to auditory streaming of frequency-modulated tones. J. Exp. Psychol. 11, 257–271.

Bartlett, E.L., Wang, X., 2005. Long-lasting modulation by stimulus context in primate auditory cortex. J. Neurophysiol. 94, 83–104.

Beauvois, M.W., Meddis, R., 1996. Computer simulation of auditory stream segregation in alternating-tone sequences. J. Acoust. Soc. Am. 99, 2270–2280.

Beauvois, M.W., Meddis, R., 1997. Time decay of auditory stream biasing. Percept. Psychophys. 59, 81–86.

Bee, M.A., Klump, G.M., 2004. Primitive auditory stream segregation: a neurophysiological study in the songbird forebrain. J. Neurophysiol. 92, 1088–1104.

Bee, M.A., 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.

Bendor, D., Wang, X., 2005. The neuronal representation of pitch in primate auditory cortex. Nature 436, 1161–1165.

Bregman, A.S., 1978. Auditory streaming is cumulative. J. Exp. Psychol.: Hum. Percept. Perf. 4, 380–387.

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

Bregman, A., Ahad, P.A., Crum, P.A.C., O'Reilly, J., 2000. Effects of time intervals and tone durations on auditory stream segregation. Percept. Psychophys. 62, 626–636.

Brosch, M., Schreiner, C.E., 1997. Time course of forward masking tuning curves in cat primary auditory cortex. J. Neurophysiol. 77, 923–943.

Butler, R.A., 1968. Effect of changes in stimulus frequency and intensity on habituation of the human vertex potential. J. Acoust. Soc. Am. 44, 945–950.

Calford, M.B., Semple, M.N., 1995. Monaural inhibition in cat auditory cortex. J Neurophysiol. 73, 1876–1891.

Carlyon, R.P., Shackleton, T.M., 1994. Comparing the fundamental frequencies of resolved and unresolved harmonics: evidence for two pitch mechanisms. J. Acoust. Soc. Am. 95, 3541–3554.

Carlyon, R.P., 2004. How the brain separates sounds. Trends Cogn. Sci. 8, 465–471.

Carlyon, R.P., Cusack, R., 2005. Effects of attention on auditory perceptual organisation. In: Itti, L., Rees, G., Tsotsos, J. (Eds.), Neurobiology of Attention. Academic Press, San Diego, pp. 317–323.

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

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

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

Deike, S., Gaschler-Markefski, B., Brechman, A., Scheich, H., 2004. Auditory stream segregation relying on timbre involves left auditory cortex. Neuroreport 15, 1511–1514.

Denham, S.L., 2001. Cortical synaptic depression and auditory perception. In: Greenberg, S., Slaney, M. (Eds.), Computational Models of Auditory Function. IOS Press, Amsterdam, pp. 281–296.

Deutsch, D., 1974. An auditory illusion. Nature 251, 307–309.

Eggermont, J.J., 1999. The magnitude and phase of temporal modulation transfer functions in cat auditory cortex. J. Neurosci. 19, 2780–2788.

Fay, R.R., 1998. Auditory stream segregation in goldfish (*Carassius auratus*). Hear. Res. 120, 69–76.

Fay, R.R., 2000. Frequency contrasts underlying auditory stream segregation in goldfish. J. Assoc. Res. Otolaryngol. 01, 120–128.

Fishman, Y.I., Reser, D.H., Arezzo, J.C., Steinschneider, M., 2001. Neural correlates of auditory stream segregation in primary auditory cortex of the awake monkey. Hear. Res. 151, 167–187.

Fishman, Y.I., Arezzo, J.C., Steinschneider, M., 2004. Auditory stream segregation in monkey auditory cortex: effects of frequency separation, presentation rate, and tone duration. J. Acoust. Soc. Am. 116, 1656–1670.

Flechsig, P., 1908. Bemerkungen über die Hoersphaere des menschlichen Gehirns. Neurol. Centralbl. 27, 50–57.

Galaburda, A., Sanides, F., 1980. Cytoarchitectonic organization of the human auditory cortex. J. Comp. Neurol. 190, 597–610.

Green, D.M., Swets, J.A., 1966. Signal Detection Theory and Psychophysics. Wiley, New York.

Griffiths, T.D., Rees, G., Rees, A., Green, G.G., Witton, C., Rowe, D., Buchel, C., Turner, R., Frackowiak, R.S., 1998. Right parietal cortex is involved in the perception of sound movement in humans. Nat. Neurosci. 1, 74–79.

Griffiths, T.D., Buchel, C., Frackowiak, R.S., Patterson, R.D., 1998. Analysis of temporal structure in sound by the human brain. Nat. Neurosci. 1, 422–427.

Grimault, N., Micheyl, C., Carlyon, R.P., Arthaud, P., Collet, L., 2000. Influence of peripheral resolvability on the perceptual segregation of harmonic complex tones differing in fundamental frequency. J. Acoust. Soc. Am. 108, 263–271.

Grimault, N., Bacon, S.P., Micheyl, C., 2002. Auditory stream segregation on the basis of amplitude-modulation rate. J. Acoust. Soc. Am. 111, 1340–1348.

Gutschalk, A., Patterson, R.D., Rupp, A., Uppenkamp, S., Scherg, M., 2002. Sustained magnetic fields reveal separate sites for sound level and temporal regularity in human auditory cortex. Neuroimage 15, 207–216.

Gutschalk, A., Patterson, R.D., Scherg, M., Uppenkamp, S., Rupp, A., 2004. Temporal dynamics of pitch in human auditory cortex. Neuroimage 22, 755–766.

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

Gutschalk, A., Melcher, J.R., Micheyl, C., Wilson, E.C., Oxenham, A.J. 2006. Neural correlates of streaming without spectral cues in human auditory cortex. Abstr. Assoc. Res. Otolaryngol. In: (29th Midwinter Research Meeting), Abstr. 811.

Hackett, T.A., Preuss, T.M., Kaas, J.H., 2001. Architectonic identification of the core region in auditory cortex of macaques, chimpanzees, and humans. J. Comp. Neurol. 441, 197–222.

Harms, M.P., Melcher, J.R., 2002. Sound repetition rate in the human auditory pathway: representations in the waveshape and amplitude of fMRI activation. J. Neurophysiol. 88, 1433–1450.

Harms, M.P., Melcher, J.R., 2003. Detection and quantification of a wide range of fMRI temporal responses using a physiologically-motivated basis set. Hum. Brain Mapp. 20, 168–183.

Harms, M.P., Guinan Jr., J.J., Sigalovsky, I.S., Melcher, J.R., 2005. Short-term sound temporal envelope characteristics determine multisecond time-patterns of activity in human auditory cortex as shown by fMRI. J. Neurophysiol. 93, 210–222.

Hartmann, W.M., Johnson, D., 1991. Stream segregation and peripheral channeling. Music Percept. 9, 155–184.

Hulse, S.H., MacDougall-Shackleton, S.A., Wisniewski, A.B., 1997. Auditory scene analysis by songbirds: stream segregation of birdsong by European starlings (*Sturnus vulgaris*). J. Comp. Psychol. 111, 3–13.

Hung, J., Jones, S.J., Vaz Pato, M., 2001. Scalp potentials to pitch change in rapid tone sequences. A correlate of sequential stream segregation. Exp. Brain. Res. 140, 56–65.

Izumi, A., 2002. Auditory stream segregation in Japanese monkeys. Cognition 82, 113–122.

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

Jones, S.J., Longe, O., Vaz Pato, M., 1998. Auditory evoked potentials to abrupt pitch and timbre change of complex tones: electrophysiological evidence of 'streaming'? Electroencephalogr. Clin. Neurophysiol. 108, 131–142.

Kanwal, J.S., Medvedev, A.V., Micheyl, C., 2003. Neurodynamics of auditory stream segregation: tracking sounds in the mustached bat's natural environment. Network: Comput. Neural Syst. 14, 413–435.

Kashino, M., Okada M., Mizutani, S., Davis, P., Kondo, H. in press. The dynamics of auditory streaming: psychophysics, neuroimaging, and modeling. In: Kollmeier, B., Klump, G., Hohmann, V, Langemann, U, Mauermann, M, Uppenkamp, S, Verhey, J. (Eds.), Hearing – from basic research to applications. Springer, Berlin.

Krumbholz, K., Patterson, R.D., Seither-Preisler, A., Lammertmann, C., Lutkenhoner, B., 2003. Neuromagnetic evidence for a pitch processing center in Heschl's gyrus. Cereb. Cortex. 13, 765–772.

Liegeois-Chauvel, C., Musolino, A., Chauvel, P., 1991. Localization of the primary auditory area in man. Brain 114, 139–151.

Liegeois-Chauvel, C., Musolino, A., Badier, J.M., Marquis, P., Chauvel, P., 1994. Evoked potentials recorded from the auditory cortex in man: evaluation and topography of the middle latency components. Electroenceph. Clin. Neurophysiol. 92, 204–214.

MacDougall-Shackleton, S.A., Hulse, S.H., Gentner, T.Q., White, W., 1998. Auditory scene analysis by European starlings (Sturnus Vulgaris): perceptual segregation of tone sequences. J. Acoust. Soc. Am. 103, 3581–3587.

McCabe, S.L., Denham, M.J., 1997. A model of auditory streaming. J. Acoust. Soc. Am. 101, 1611–1621.

Micheyl, C., Tian, B., Carlyon, B.P., Rauschecker, J.P., 2005. Perceptual organization of sound sequences in the auditory cortex of awake macaques. Neuron 48, 139–148.

Micheyl, C., Shamma, S., Oxenham, A.J. in press. Hearing out repeating elements in randomly varying multitone sequences: a case of streaming. In: Kollmeier, B., Klump, G., Hohmann, V, Langemann, U, Mauermann, M, Uppenkamp, S, Verhey, J. (Eds.), Hearing – from basic research to applications. Springer, Berlin.

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

Moore, B.C.J., Gockel, H., 2002. Factors influencing sequential stream segregation. Acta Acustica 88, 320–333.

Morosan, P., Rademacher, J., Schleicher, A., Amunts, K., Schormann, T., Zilles, K., 2001. Human primary auditory cortex: cytoarchitectonic subdivisions and mapping into a spatial reference system. Neuroimage 13, 684–701.

Näätänen, R., Sams, M., Alho, K., Paavilainen, P., Reinikainen, K., Sokolov, E.N., 1988. Frequency and location specificity of the human vertex N1 wave. Electroenceph. Clin. Neurophysiol. 69, 523–531.

Näätänen, R., Tervaniemi, M., Sussman, E., Paavilainen, P., Winkler, I., 2001. Primitive intelligence in the auditory cortex. Trends Neurosci. 24, 283–288.

Pantev, C., Bertrand, O., Eulitz, C., Verkindt, C., Hampson, S., Schuierer, G., Elbert, T., 1995. Specific tonotopic organization of different areas of the human auditory cortex revealed by simultaneous magnetic and electric recordings. Electroenceph. Clin. Neurophysiol. 94, 26–40.

Parker, A.J., Newsome, W.T., 1998. Sense and the single neuron: probing the physiology of perception. Ann. Rev. Neurosci. 21, 227–277.

Patterson, R.D., Uppenkamp, S., Johnsrude, I.S., Griffiths, T.D., 2002. The processing of temporal pitch and melody information in auditory cortex. Neuron 36, 767–776.

Penagos, H., Melcher, J.R., Oxenham, A.J., 2004. A neural representation of pitch salience in nonprimary human auditory cortex revealed with functional magnetic resonance imaging. J. Neurosci. 24, 6810–6815.

Phillips, D.P., Irvine, D.R., 1981. Responses of single neurons in physiologically defined primary auditory cortex (AI) of the cat: frequency tuning and responses to intensity. J. Neurophysiol. 45, 48–58.

Picton, T.W., Woods, D.L., Proulx, G.B., 1978. Human auditory sustained potentials. II. Stimulus relationships. Electroenceph. Clin. Neurophysiol. 45, 198–210.

Pressnitzer, D., Hupe, J.M., 2006. Temporal dynamics of auditory and visual bistability reveal common principles of perceptual organization. Curr. Biol. 11, 1351–1357.

Rauschecker, J.P., Tian, B., 2004. Processing of band-passed noise in the lateral auditory belt cortex of the rhesus monkey. J. Neurophysiol. 91, 2578–2589.

Rauschecker, J.P., Tian, B., Hauser, M., 1995. Processing of complex sounds in the macaque nonprimary auditory cortex. Science 268, 111–114.

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

Scherg, M., 1990. Fundamentals of dipole source analysis. In: Grandori, F., Hoke, M., Romani, G.L. (Eds.), Auditory Evoked Magnetic Fields and Electric Potentials. In: Advances in Audiology, vol. 6. Karger, Basel, pp. 40–69.

Snyder, J.S., Alain, C. submitted for publication. Neurophysiological mechanisms of auditory stream segregation.

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

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

Tian, B., Rauschecker, J.P., 2004. Processing of frequency-modulated sounds in the lateral auditory belt cortex of the rhesus monkey. J. Neurophysiol. 92, 2993–3013.

Ulanovsky, N., Las, L., Nelken, I., 2003. Processing of low-probability sounds by cortical neurons. Nat. Neurosci. 6, 391–398.

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

van Noorden, L.P.A.S., 1975. Temporal Coherence in the Perception of Tone Sequences. University of Technology, Eindhoven.

Vliegen, J., Oxenham, A.J., 1999. Sequential stream segregation in the absence of spectral cues. J. Acoust. Soc. Am. 105, 339–346.

Vliegen, J., Moore, B.C.J., Oxenham, A.J., 1999. The role of spectral and periodicity cues in auditory stream segregation, measured using a temporal discrimination task. J. Acoust. Soc. Am. 106, 938–945.

Wehr, M., Zador, A.M., 2005. Synaptic mechanisms of forward suppression in rat auditory cortex. Neuron 47, 437–445.

Wilson, E.C., Melcher, J., Micheyl, C., Gutschalk, A., Oxenham, A. 2005. Auditory streaming in humans: possible fMRI correlates in auditory cortex. Abstr. Assoc. Res. Otolaryngol. (28th Midwinter Research Meeting), Abstr. 466.

Yin, P., Ma, L., Elhilali M., Fritz, J., Shamma, S. in press. Primary auditory cortical responses while attending to different streams. In: Kollmeier, B., Klump, G., Hohmann, V, Langemann, U, Mauermann, M, Uppenkamp, S, and Verhey, J., (Eds)., Hearing – from basic research to applications. Springer, Berlin.

Yvert, B., Crouzeix, A., Bertrand, O., Seither-Preisler, A., Pantev, C., 2001. Multiple supratemporal sources of magnetic and electric auditory middle latency components in humans. Cereb. Cortex 11, 411–423.