Research paper

Active stream segregation specifically involves the left human auditory cortex

Susann Deike *, Henning Scheich, André Brechmann

Leibniz Institute for Neurobiology, Brenneckestr. 6, 39118 Magdeburg, Germany

ARTICLE INFO

ABSTRACT

An important aspect of auditory scene analysis is the sequential grouping of similar sounds into one "auditory stream" while keeping competing streams separate. In the present low-noise fMRI study we presented sequences of alternating high-pitch (A) and low-pitch (B) complex harmonic tones using acoustic parameters that allow the perception of either two separate streams or one alternating stream. However, the subjects were instructed to actively and continuously segregate the A from the B stream. This was controlled by the additional instruction to listen for rare level deviants only in the low-pitch stream.

Compared to the control condition in which only one non-separable stream was presented the active segregation of the A from the B stream led to a selective increase of activation in the left auditory cortex (AC). Together with a similar finding from a previous study using a different acoustic cue for streaming, namely timbre, this suggests that the left auditory cortex plays a dominant role in active sequential stream segregation. However, we found cue differences within the left AC: Whereas in the posterior areas, including the planum temporale, activation increased for both acoustic cues, the anterior areas, including Heschl's gyrus, are only involved in stream segregation based on pitch.

© 2010 Elsevier B.V. All rights reserved.

1. Introduction

In recent years the search for the neural mechanisms underlying "auditory scene analysis" (Bregman, 1990) has become an important topic in neuroscience (for reviews see Carlyon, 2004; Denham and Winkler, 2006; Micheyl et al., 2007; Snyder and Alain, 2007). The basic question is how the auditory system segregates a mixture of competing acoustic sequences into distinct meaningful auditory objects or "auditory streams" referring to certain sound sources (Bregman, 1990). Experimentally, sequential stream segregation is mostly studied by using sequences of repeating tones presented in an ABA_ or ABAB design. In most studies, A and B pure tones with differing frequencies are used. Depending on the frequency difference (ΔF) and the presentation rate of A and B tones, three different perceptual domains can be distinguished by determining the fission and coherence boundary (van Noorden, 1975). Below the fission boundary, only a single alternating stream is perceived; beyond the coherence boundary, two segregated streams

are perceived; and between these boundaries, both of these percepts are possible (ambiguous domain).

Results from electrophysiological recording studies in animals have led to a hypothesis on the neuronal mechanism that explains the percept of two separate streams (Bee and Klump, 2004, 2005; Fishman et al., 2001, 2004; Kanwal et al., 2003; Micheyl et al., 2005). It is suggested that frequency selectivity of tonotopically organized neurons in primary auditory cortex fields in combination with physiological forward suppression leads to separate representations of A and B tones. The differential suppression of non-best frequency tones was suggested to be the neuronal basis for the percept of two separate A and B streams. However, recent studies have revealed that the subcortical auditory pathway also plays a role in auditory stream formation (Kondo and Kashino, 2009; Pressnitzer et al., 2008) and besides tonotopic neural response separation, temporal coherence may have an influence as well (Elhilali et al., 2009).

Within the ambiguous perceptual domain of streaming, the perceptual organization depends on the attentional set of the subject and the instructions given as first described by van Noorden (1975) (for review see Moore and Gockel, 2002). This strongly suggests additional top–down mechanisms involved in stream segregation in the ambiguous domain that are independent of the physical stimulus.

A number of human imaging studies investigated stream segregation, including perceptually ambiguous tone sequences: EEG and MEG studies using mismatch negativity (MMN) paradigms

introduced deviants into sequences of ABAB tones which were easier to detect when the perceptual organization of two separate streams was present. However, MMN responses to the deviants were only observed when the subjects were cued to perceive two separate streams either by explicit instruction (Sussman et al., 1998), by initial priming for a segregated organization (Sussman and Steinschneider, 2006), or by visual stimuli synchronized to the tones of one separate stream (Rahne et al., 2007, 2008).

Further studies systematically varied the frequency separation between the pure tones of ABA_ sequences and found increases of auditory evoked potentials/fields to the B tones corresponding to larger frequency separations between A and B pure tones (P1–N1–P2 and N1c in Snyder et al. (2006) and P1m–N1m in Gutschalk et al. (2005, first experiment)). This enhancement also correlated with behavioral reports of stream segregation. Similarly, Gutschalk et al. (2007) described an increase of the P1m component for the A tone with increasing separation in fundamental frequencies (f_0) between harmonic complex tones in repeating ABBB_ sequences where A and B denote tones with different f_0. In the second experiment of the earlier MEG study, Gutschalk et al. (2005) directly compared magnetic fields evoked by B tones of a certain perceptually ambiguous ABA_ sequence when the subjects perceived two separate streams and when the subjects perceived one integrated stream. They found a larger N1m amplitude during the perceptual organization of two separate streams both when the subjects followed the A or B tones and a larger P1 amplitude when the subjects followed the B tones. Gutschalk et al. (2005, 2007) and Snyder et al. (2006) suggested that reduced forward suppression may be the neuronal mechanism underlying the enhanced long latency components/fields during the perception of two segregated streams. The dipoles of these components/fields were located in the non-primary auditory cortex.

This extends findings from electrophysiological recordings in animals and suggests additional stream segregation mechanisms beyond the primary auditory cortex.

An involvement of non-primary auditory cortex areas in stream segregation was also shown in two fMRI studies (Gutschalk et al., 2007; Wilson et al., 2007). Similar to the approach by Snyder et al. (2006) and Gutschalk et al. (2005, 2007, first experiment) they used different frequency/f_0 separations between A and B tones in repeating ABBB_ and ABAB sequences, respectively, and also found increasing fMRI activation in Heschl's gyrus and the planum temporale with increasing frequency/f_0 separation between the A and B tones.

In an fMRI study Cusack (2005) used ambiguous sequences of ABA_ tones and compared the activation when the subjects heard two separate streams or one alternating stream. Cusack did not find activation differences in the auditory cortex, but a difference in the right intraparietal sulcus. In a recent fMRI study (Rahne et al., 2008), also no differences in auditory cortex activation were observe when using an ambiguous sound sequence (ABAB) whose perception was differentially biased towards either an integrated or segregated percept by appropriate visual stimuli. A third fMRI study employing an ambiguous sound sequence, used timbre as acoustic parameter for segregation of ABAB sounds produced by an organ or a trumpet (Deike et al., 2004). They instructed the subjects to segregate the A- and B-timbre streams by continuously following the sound of a specific timbre. This led to an increase of fMRI activation in non-primary auditory cortex of the left hemisphere compared to a non-separable control stream.

In summary, a number of imaging studies suggest that additional mechanisms are involved beyond the one suggested for stream segregation based on pitch first put forward by Fishman (Fishman et al., 2001, 2004). The first reason is that in contrast to the electrophysiological recordings in animals, that were restricted to the primary auditory cortex, the above mentioned human imag-

ing studies showed an involvement of secondary auditory areas. Secondly, when using the same ambiguous tone sequence, additional mechanisms need to be involved to evoke one or the other perceptual organization and this may be reflected in the activation of auditory cortex. As this is independent of the physical stimulus properties we refer to such mechanisms as "top–down". In previous streaming studies such top–down influences on the activation in auditory cortex were initialized by instructing the subjects to select one or the other perceptual organization either by priming (Sussman and Steinschneider, 2006), synchronized visual stimuli (Rahne et al., 2007) or explicit instructions (Deike et al., 2004; Sussman et al., 1998).

The aim of the current study was to investigate top–down mechanisms that are required to maintain a segregated percept of an ambiguous tone sequence instead of freely switching between the two possible perceptual organizations. Therefore, we instructed our subjects to actively and continuously segregate auditory streams on the basis of pitch differences. Our second goal was to localize and differentiate these task-specific effects within the auditory territories previously defined by Brechmann et al. (2002) and beyond that in more detail. The third aim was to clarify the specific role of the left hemisphere in stream segregation that was suggested in the study by Deike et al. (2004) using timbre as the acoustic parameter for segregation.

2. Materials and methods

2.1. Subjects

Twenty right-handed (Edinburgh Handedness Inventory) normal hearing subjects (seven male, 13 female, from 20 to 35 years old) participated in the fMRI study, 11 in experiment I and nine in experiment II. All participants showed a language laterality toward the left hemisphere tested as described in Bethmann et al. (2006). Six participants from experiment II and 11 additional listeners took part in psychophysical measurements. The subjects gave written informed consent to the study which was approved by the Ethics Committee of the University of Magdeburg.

2.2. Stimuli and task

2.2.1. Psychophysics

Psychophysical measurements were performed to check that the stimulus used in the fMRI experiments evoked ambiguous percepts. We measured the proportion of time that these sequences were perceived as two separate streams. The stimuli were digitally synthesized (Cool Edit 2000) harmonic tones comprising 10 harmonics of equal amplitude, lasting 200 ms including linear ramps of 10 ms at the beginning and end. The fundamental frequencies were chosen from two ranges corresponding to 4 tones of the C4 octave (261, 293, 329, 349 Hz) and the C5 octave (523, 587, 659, 698 Hz). High-pitch (A) and low-pitch (B) complex harmonic tones were presented alternately in an ABAB design with a presentation rate of 5 Hz. Within a frequency range (C4 or C5 octave), fundamental frequencies varied randomly to prevent habituation effects leading to reduced overall activation in auditory cortex. The stimuli were arranged in nine blocks of 27 s which were alternately presented with silence blocks lasting 5 s. Subjects were instructed to indicate the currently prevailing sound organization (one *integrated* stream or two *segregated* streams) by pressing the left and right mouse buttons, respectively. The experiment was performed in a quiet booth on a PC using the software Presentation (Neurobehavioral Systems, Inc., San Francisco) and headphones (Sennheiser, HD 465).

The total time during which the listener reported hearing the stimulus sequence as two separate streams was divided by the se-

quence duration minus the time of the first response. The result was regarded as an estimate of the proportion of time that the sequence was heard as segregated. Across the group the *segregated* percept occurred 64.41% (±3.00%) of the time. According to the definition by van Noorden (1975), the stimulus sequences were thus perceptually ambiguous.

2.2.2. fMRI

fMRI measurements were made to test the effect of active stream segregation on the BOLD response. In the segregation condition the same ambiguous stimulus sequences as in the psychophysical measurements were used and the subjects were instructed to actively segregate the alternating harmonic tones into two separate streams and follow the low tones. In order to additionally encourage the participants to segregate, the subjects had to detect occasional "deviant" tones in the low-pitch stream. These tones had a level 4 dB higher than the "standard" tones and occurred at random with a mean frequency of occurrence of 15% relative to the number of total sounds in that stream. The same level deviants with the same mean frequency of occurrence were distributed in the high-pitch stream, serving as distractors.

In the control condition, subjects had to perform the same task (level deviants +4 dB SPL, 15% mean frequency of occurrence) but this time on a single, perceptually inseparable stream of harmonic tones with the same duration and the same fundamental frequencies as the low-pitch tones (C4 octave) of the stream segregation condition. Furthermore, the control stimuli were adjusted to the test stimuli with respect to the frequency bandwidth: The number of harmonic components were increased to seventeen and thus roughly covered the long-term spectrum of both the low- and the high-pitch tones of the segregation condition. The harmonic tones were presented with randomly varying fundamental frequencies and with either a presentation rate of 2.5 Hz, which corresponds to the perceived rate of one separated stream in the segregation condition (experiment I), or with 5 Hz, which corresponds to the physical rate of the tone sequence in the segregation condition (experiment II). Ten subjects were asked whether they were able to segregate the tone sequence of the control condition but none reported to be able to do so.

Stimuli were arranged in blocks of 27 s duration and each covered one of the two conditions: segregation ($n = 9$) and control ($n = 9$). Between the randomly distributed stimulus blocks, silence blocks of the same duration served as the resting condition.

For stimulus presentation and recording of behavioral responses the software Presentation (Neurobehavioral Systems, Inc., San Francisco) was used. The stimuli were presented via fMRI-compatible electrodynamic headphones integrated into ear-muffs for reduction of residual background scanner noise (Baumgart et al., 1998). The presentation level of the stimuli was adjusted to a comfortable level of about 70 dB SPL at which the level deviants were clearly audible.

2.3. The fMRI data acquisition and preprocessing

Subjects were scanned in a 3 T BRUKER 30/60 head scanner equipped with a birdcage coil and an asymmetric gradient system. Three contiguous slices of 6 mm thickness each were oriented parallel to the Sylvian fissure covering the superior temporal gyrus of both hemispheres. Functional volumes were collected using a low-noise FLASH-based gradient echo sequence (TE/TR/flip = 35 ms/140.7 ms/15°, matrix 64×64, FOV 18×18 cm²). A long gradient rise time (2500 µs) reduced the scanner noise to a level of 54 dB (A) peak amplitude (www.ifn-magdeburg.de/assets/files/special_labs/non-invasive_brain_imaging/methods/3/1/3-1-fmri-scanner-ramp.wav). With this setting, we think that the scanner noise had only minor effects on the listeners' percept and that the ABAB tone sequence was still perceived as ambiguous.

A total fMRI experiment comprised 111 volumes scanned in 19 min and 36 s. In order to obtain anatomical landmarks, functional measurements were followed by the acquisition of a high resolution T1-weighted volume (MDEFT). The subject's head was fixed with a vacuum cushion with attached ear-muffs containing the headphones. The subjects were instructed to keep their eyes closed throughout the entire fMRI session.

Each functional dataset was subjected to a quality check: First, the subject's head motion during the functional measurement was detected using the AIR package (Woods et al., 1998). Data with continuous head motion greater than one voxel in at least one direction were excluded from further analysis. Furthermore, the mean gray value of voxels covering the temporal lobe defined in two slices was computed for all functional volumes. Single images with gray value deviations of more than 2.5 % compared to the mean gray value of all volumes were excluded from further analysis. In case of exclusion of more than two images of one stimulus condition, the whole functional dataset was discarded. In addition, a 2D motion correction was applied.

The functional datasets were analyzed with the software package KHORFu (Gaschler et al., 1996). The matrix size of 64×64 was increased to 128×128 by pixel replication followed by in-plane smoothing with a Gaussian filter [FWHM = 2 pixel (2.8 mm), Kernel width = 5 pixel (7 mm)]. For each subsequent scan of the same slice, the mean intensity was computed and then scaled to the mean slice intensity average over all time points. Afterwards each voxel time series was temporally smoothed using a moving average filter with a kernel width of two time points.

2.4. Individual regions of interest [ROI] analysis

We defined the regions of interest in the auditory cortex as described in Brechmann et al. (2002). This approach has proven useful for regional comparison because a functional parcellation of human auditory cortex is not yet available and grand averages of brain transforms tend to blur and mislocalize activations in the superior temporal lobe due to large interindividual anatomical variability (Leonard et al., 1998; Penhune et al., 1996).

For each subject functional activation in each slice was analyzed by correlation analysis to obtain a statistical parametric map. A trapezoid function served as the correlation vector, roughly modeling the expected BOLD response: thereby the first image of each stimulus and silence block was set to half-maximum values. Pearson's correlation analysis tested either the stream segregation or the control conditions versus the resting condition. Voxels were attributed as activated if the significance level was below 0.05 and if they belonged to a cluster of at least eight contiguous voxels. These clusters in each slice were attributed to one of four territories TA, T1, T2, and T3 defined as previously described (Brechmann et al., 2002; Scheich et al., 1998). For each territory in each subject and each condition, the total number of activated voxels was multiplied by their average relative BOLD signal intensity, resulting in intensity weighted volumes (IWV). The IWV of the entire AC and each territory was subjected to a repeated-measure analysis of variance (ANOVA), testing the within-subject factor of the *perceptual organization* (segregation, control) and the between-subject factor of the *presentation rate* (perceptually (experiment I) or physically (experiment II) identical to the segregation condition). We applied a Bonferroni–Holm correction ($\alpha_{global} = 0.05$) for both the ANOVA on the IWV of the whole left and right auditory cortex (2 tests) and the ANOVA of the IWV of the individual territories (8 tests).

2.5. Further subdivision of the territories T1 and T2

The distribution of *p*-values of the statistical parametric map along the long axis of Heschl's gyrus revealed distinct clusters

within T1 and T2. This was used to further subdivide T1 and T2 into a medial and lateral area using minima of *p*-values between these clusters (Fig. 4A–C).

Pearson's correlation analysis was performed within the subterritories to test each condition versus rest. The average time course of the BOLD response of these significantly activated voxels ($p < 0.05$, cluster size >7) was analyzed for each subterritory using the linear contrast between the stream segregation and the control condition (GLM-analysis). Additionally, the peak in the average time course of each condition was tested across subjects for differences between the medial and lateral subdivisions of T1 and T2 (two-tailed *t*-test, $p < 0.05$).

2.6. Statistical analysis of task performance

Behavioral responses of subjects were recorded with respect to the number of hits, false alarms, misses and correct rejections. In the segregation condition the correct answers were reactions to level deviants in the target stream; false alarms were reactions to level deviants in the non-target stream as well as reactions to non-deviants in both streams; missing answers were non-reactions to level deviants in the target stream and correct rejections were non-reactions to non-deviants in both streams and deviants in the non-target stream. Button presses were counted as hit when they occurred in the response window between 300 and 1000 ms after the onset of a target stimulus. Any other response was counted as a false alarm. From these measures we calculated the sensitivity index d′ (Swets et al., 1961) for each condition and for each subject. To determine whether the subjects were able to perform the task and thus to segregate the two streams, we subjected the number of hits, false alarms, misses and correct rejections to a chi-square test (one-sided, $u > 6.63$). We chose a significance level of $p < 0.01$ to make sure that the subjects were indeed able to perform the task. Subjects who did not reach this criterion were excluded from further analysis (one case).

Similarly to the statistical analysis of fMRI activation, the sensitivity indices d′ were subjected to a repeated-measure ANOVA testing for the within-subject effect of the perceptual *organization* (segregation, control) and the between-subject effect of the *presentation rate* (presentation rate of the control condition: perceptually (experiment I) or physically (experiment II) identical to the segregation condition).

3. Results

3.1. Task performance

The repeated-measure ANOVA of sensitivity index d′ revealed a significant main effect of the within-subject factor *perceptual organization* (segregation, control) with a lower d′ in the segregation condition compared to both control conditions [$F(1, 16) = 9.5$, ($p = 0.007$)]. Furthermore, the ANOVA revealed a significantly higher overall task performance in experiment II using the same physical presentation rate in the segregation and control condition than in experiment I using the same perceptual presentation rate in the segregation and control condition (Main effect of *presentation rate* [$F(1, 16) = 12.3$, ($p = 0.003$)] with no interaction between *perceptual organization* and *presentation rate* [$F(1, 16) = 0.9$, ($p = 0.34$)]). The average performance of the subjects in the detection task is given in Fig. 1.

3.2. Individual ROI analysis

The repeated-measure ANOVA revealed a significant main effect of the *perceptual organization* (segregation, control) on fMRI activa-
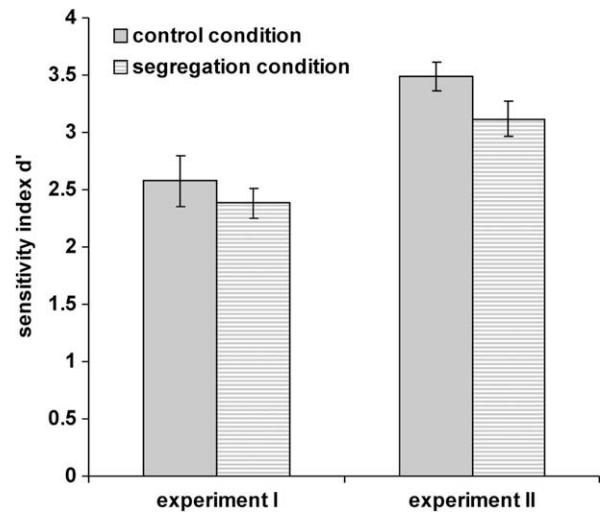

Fig. 1. Average performance of the subjects in the detection task. The sensitivity index d′ in the segregation condition was significantly lower than in the control condition irrespective of the presentation rate of the control condition. Furthermore, the overall performance of detecting level deviants was higher in experiment II than in the experiment I for both the segregation and the control conditions.

tion (IWV) for the entire left AC [$F(1, 17) = 29.3$, ($p = 0.00006$)] and all left auditory territories [TA: $F(1, 17) = 16.8$, ($p = 0.001$); T1: $F(1, 17) = 11.5$, ($p = 0.004$); T2: $F(1, 17) = 10.4$, ($p = 0.005$); and T3: $F(1, 17) = 8.7$, ($p = 0.01$)], with stronger activation (IWV) in the segregation condition compared to the control condition. All described results are still significant after the Bonferroni–Holm correction. This is illustrated in Fig. 3 as the difference between the activation of the segregation and the control condition for comparison with the results of our previous study using timbre cues for segregation (Deike et al., 2004). In the right AC, we found no effect of the *perceptual organization* on either the global or for the territorial activation [$F(1, 17) < 3.3$, ($p > 0.09$)]. There was no significant main effect of *presentation rate* and no significant interaction between the factors *perceptual organization* and *presentation rate*.

The main result of the statistical analysis was an equal effect of the *perceptual organization* on activation (IWV) and was thus independent of the *presentation rate* used in the control condition (experiment I: 2.5 Hz; experiment II: 5 Hz). In addition, this effect was restricted to the left AC. An example of the activation pattern in the auditory cortex and a schematic parcellation into territories is given in Fig. 2A and B for an individual subject.

3.3. Subdivision of the territories T1 and T2

In T2, all nineteen volunteers showed two clearly separable foci of activation. In T1, two separable foci were found in seventeen of nineteen volunteers (Fig. 4C). The remaining two volunteers showed only one focus of activation located on the lateral aspect of T1 at the selected significance level (Fig. 4B).

To test whether the separate clusters of activation within T1 and T2 contributed equally to the streaming-specific effects found for the entire territories T1 and T2 of the left hemisphere, we subdivided T1 and T2 into a medial and a lateral part. In some subjects we found a third focus of activation in T1 located at the base of Heschl's gyrus (Fig. 4A). However, since only three of nineteen subjects clearly showed this separate cluster, we subsumed this cluster into the medial part of T1 (Fig. 4A).

GLM-analysis of the BOLD response of significantly activated voxels revealed a stronger activation in stream segregation compared to the control condition in all subregions (lateral T1:
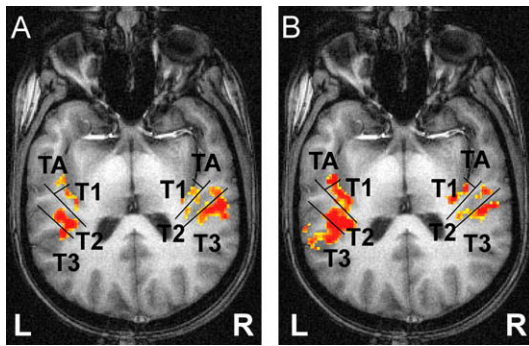
Fig. 2. Activation in auditory cortex territories. Pattern of significant activations ($p < 0.05$) in one slice of an individual subject in the control condition (A) and in the stream segregation condition (B). Black lines schematically indicate the borders between territories. Significantly activated voxel ($p < 0.05$) are color coded from yellow to red on a logarithmic scale.
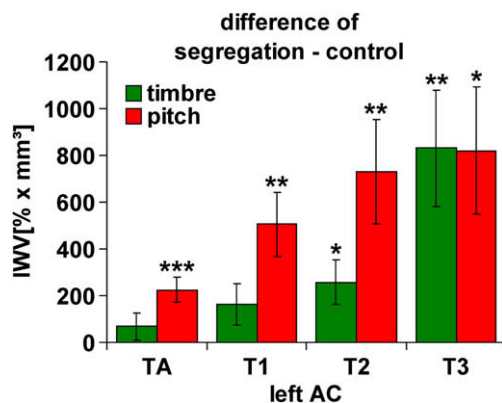

Fig. 3. Activation differences between conditions on the level of single territories within the left hemisphere illustrating the significant increase of activation in the segregation condition compared to the control condition. The data from the current study were plotted in red and for comparison the data from the previous timbre study were included and plotted in green. Note that in both studies a significant increase of activation in the segregation condition compared to the control condition was observed in the territories T2 and T3, suggesting a general function of these territories in stream segregation tasks. In contrast, the anterior areas TA and T1 are involved only in stream segregation relying on pitch.

$t = 3.9$, $p = 0.00009$; medial T1: $t = 3.9$, $p = 0.0001$; lateral T2: $t = 4.3$, $p = 0.00002$; medial T2: $t = 2.6$, $p = 0.01$).

A comparison (two-tailed t-test) of the peak in the average time course between the medial and lateral parts of T1 and T2 revealed for both conditions a significantly larger BOLD response in the lateral parts (T1: segregation $t = 2.5$, $p = 0.02$; control: $t = 3.2$, $p = 0.006$; T2: segregation $t = 5.9$, $p = 0.00002$; control: $t = 5.7$, $p = 0.00003$). The average time courses of the subterritories are shown in Fig. 4D–G.

4. Discussion

4.1. Selective involvement of the left auditory cortex

The main result of this study is the selective involvement of the left auditory cortex when the subjects had to actively and continuously segregate the A from the B stream, using pitch differences as an acoustic cue. From a behavioral point of view, we first showed that ABAB tone sequences were perceptually ambiguous. Second, we made sure that the subjects constantly segregated the A from the B stream by introducing the task of detecting small changes in stimulus level.

A left auditory cortex effect was already shown in a similar stream segregation task when subjects had to segregate ABAB tone series according to differences in timbre (Deike et al., 2004). The left auditory cortex effect in both studies is especially remarkable because both the processing of pitch, especially changes in pitch, and the processing of timbre is mainly attributed to the right hemisphere (e.g. Behne et al., 2005; Brechmann and Scheich, 2005; Hall et al., 2005; Menon et al., 2002; Patterson et al., 2002; Platel et al., 1997; Warren and Griffiths, 2003; Zatorre et al., 2002).

Other studies analyzing the neural correlates of stream segregation on the basis of pitch differences either did not observe activation differences between the left and right auditory cortex activation (Cusack, 2005; Gutschalk et al., 2005, 2007; Wilson et al., 2007) or found a larger N1c amplitude in the right compared to the left hemisphere (Snyder et al., 2006). However, the subjects in these studies were free to switch their perceptual organization and were not instructed to actively segregate the A from the B stream during the entire sequence.

We suggest that the instruction for stream segregation in combination with the detection task motivated the subjects to actively and continuously segregate the A from the B tones and that it prevented them from switching to integrate the tones into one stream. There are even some arguments that the listening task biased towards segregation. When asked about their strategy to solve the task, the subjects did not report that they attended to all tones of increased level but only responded to the low-pitch level deviants. This seems plausible because the strategy to select the low-pitch stream is the most economic way to solve the target detection within the low-pitch stream. The good task performance with only a few button presses to the deviants in the high-pitch stream indicates that the subjects indeed focused on the low-pitch stream. This interpretation is supported by Micheyl et al. (2004) who showed that the performance in a within-stream benefits from a segregated perceptual organization of ABA_ sequences. We also suggest that this process of intentional segregation that we refer to as top–down process, has led to the increase in activity in the left auditory cortex. In a previous study, Deike et al. (2004) used the same approach and made a similar observation of an increase in activation in the left but not the right auditory cortex when the subjects were instructed to actively and continuously segregate A and B tones differing in timbre. Therefore, we think that the involvement of the left auditory cortex is independent of the main cue used for stream segregation and that the reason must have to do with hemispheric differences in sound processing other than the processing of pitch and timbre per se. Bregman (1990) emphasized the sequential nature of stream segregation in his statement that tones which are similar to one another (e.g. with respect to pitch, timbre, and spatial location) must be sequentially grouped in preference to sounds that follow one another. Consequently, we argue that the demand on sequential processing during active and continuous stream segregation was responsible for the selective increase in activation in the left auditory cortex. This hypothesis was already formulated in the previous study by Deike et al. (2004), but the present results provide further evidence that the stream segregation specific effect in the left auditory cortex is independent of the cue which is used for segregation and thus not due to the processing of acoustic features like the timbre or pitch of harmonic tones.

At this point we would like to mention the possibility that the difference in spectral extent of the high- and low-pitch tones may cause different timbre that could have been used as a cue for segregation. However, none of the subjects reported that the harmonic tones differ in timbre, and that they used timbre rather than pitch for segregation. This seems plausible because the differences in timbre are rather subtle and difficult to perceive if the pitch changes (Krumhansl and Iverson, 1992). Furthermore, Singh
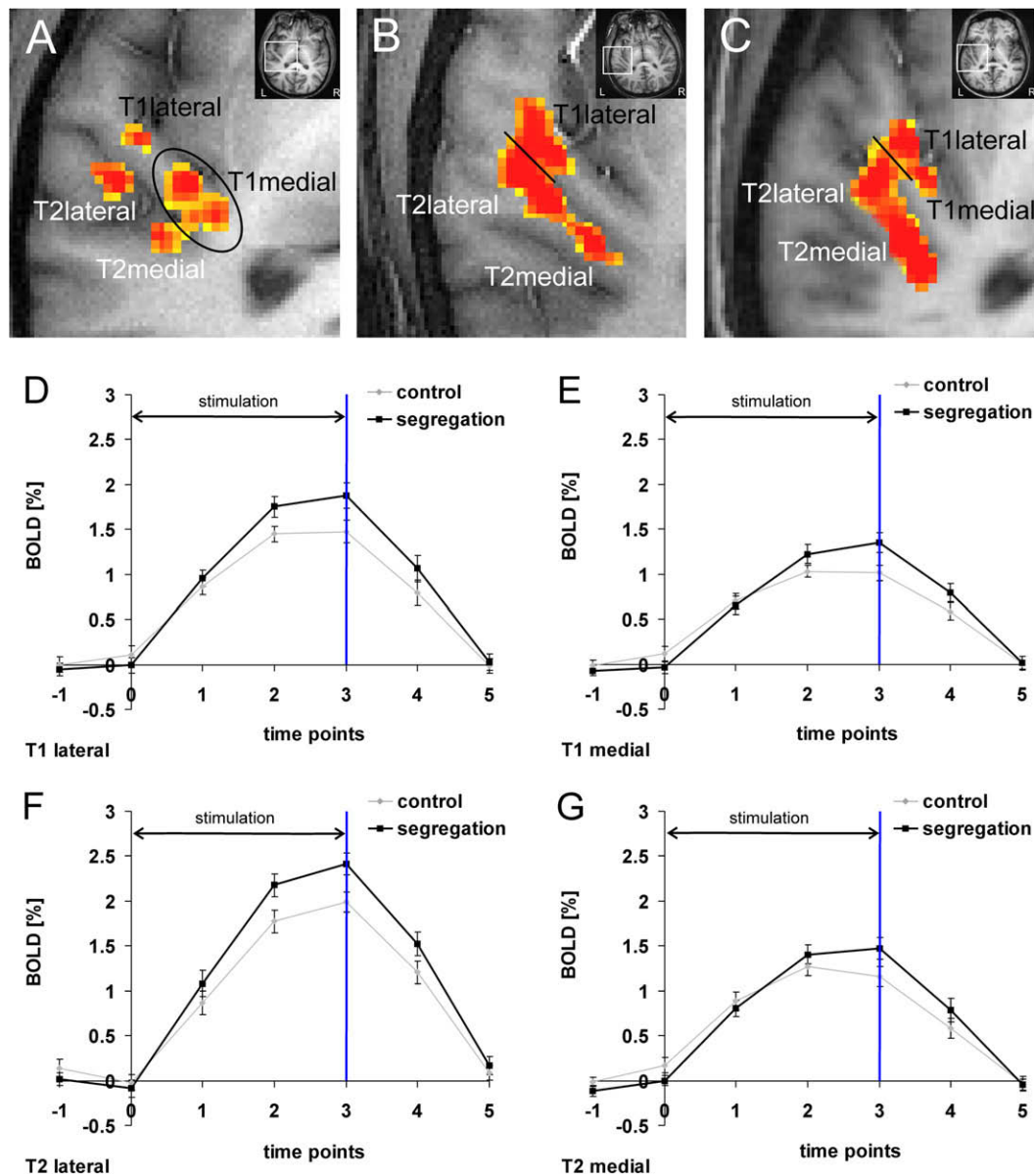
Fig. 4. Subdivisions of the territories T1 and T2. The columns (A–C) show clusters of activation in the left territories T1 (on Heschl's gyrus) and T2 (centered on Heschls sulcus) of single subjects to be used for further subdivision into a medial and lateral part and the original slices from which the image details were derived. Significantly activated voxel ($p < 0.05$) are color coded from yellow to red on a logarithmic scale. Note that in (A) the activation in the medial aspect of T1 could be further subdivided into two separate clusters. In (B), an example for only one focus in T1 is shown; in (C), two foci can be separated in T1. The columns (D–G) show the average time course of the BOLD responses in the segregation and the control condition for the lateral (D and F) and the medial (E and G) parts of T1 and T2. In all subterritories the BOLD response is significantly larger in the segregation condition compared to the control condition. The peak in the average time course of both conditions is significantly larger in the lateral aspects of T1 and T2 than in the medial ones.

(1987) suggested that pitch and timbre compete in the formation of auditory streams and that there are trade-offs between these two attributes. He described for pitch intervals greater than or equal to an octave as used here that the tendency to group by timbre decreases. Therefore it seems very likely that the subjects indeed used the most salient parameter for segregation, i.e. pitch.

There are several studies supporting the notion of the left hemisphere's specific involvement in sequential processing. In their recently published animal model of lateralized auditory cortex functions, Wetzel et al. (2008) suggested the specific role of the left auditory cortex in tasks requiring segmental comparison. In humans, Liegeois-Chauvel et al. (1999) showed that intracerebral evoked potentials reflect sequential processing of different components of voiced and voiceless syllables only in left Heschl's gyrus and left planum temporale. A right ear advantage, i.e. left hemi-

sphere dominance, for temporal order perception was found using dichotic listening in patients (Efron, 1963b) and normal subjects (Efron, 1963a; Halperin et al., 1973). Using fMRI, Binder et al. (1996) revealed left planum temporale activation when participants actively listened to sequences of high (750 Hz) and low (500 Hz) tones and responded to sequences containing two high tones.

Furthermore, there is a formal similarity of sequential stream segregation to working memory tasks in that temporally spaced sounds have to be sequentially processed. It was shown by fMRI that a working memory task requiring the matching of each frequency-modulated tone (FM) with the FM tone 2-back in the sequence (2-back task) compared to the identification of a reference FM which was learned previously to the presentation of the test sequence (0-back task) exclusively involved the left planum temporale (Brechmann et al., 2007).

The behavioral data also suggest that the level of task difficulty might explain activation differences between conditions. There is evidence in the literature that task performance influences the strength of activation in regions that are specifically involved in task processing, resulting in either an increase (Binder et al., 2004; Brechmann et al., 2007; Tregellas et al., 2006) or a decrease (Brechmann and Scheich, 2005) in activation with better task performance. In the current study, the task performance measured by the sensitivity index d' was lower in the segregation condition compared to the control condition. In both conditions the subjects had to detect randomly distributed level deviants; however, this target detection had an additional load, i.e. the instruction to separate the high- from the low-pitch stream in the segregation condition. Thus, the behavioral and the activation data converge in an effect of stream segregation indicated behaviorally by a decrease in task performance and functionally by an increase in activation.

4.2. Streaming effects in territories and subterritories of the left auditory cortex

On the level of territories and subterritories within the left auditory cortex we observed effects which were either independent or dependent of the cue used for segregation.

In both the current and the previous study (Deike et al., 2004), the active stream segregation yielded an increase of activation in the secondary areas T2 and T3 of the left auditory cortex, whereas this effect was additionally observed in the territories TA and T1 only in the current pitch study. This suggests that T2 and especially T3 cover functional fields that are involved in stream segregation tasks independent of the cue used for segregation. In their model of auditory processing, Griffiths and Warren (2002) propose a functional role for the planum temporale in auditory object representation. The increase of activation in the territory T3 on planum temporale and thus its independence of the cue used for segregation of auditory streams confirm this hypothesis on higher order representations.

On the other hand, the territories TA and T1 seem to be specifically involved in stream segregation relying on pitch. Previous studies hypothesized that the activation of the territory TA (formerly T1a) might be related to foreground–background decomposition tasks (Brechmann et al., 2002; Scheich et al., 1998), which involve the simultaneous but also sequential grouping of sound elements. We found that the activation in the territory TA depends on the acoustic cue used for stream segregation. Why sequential streaming relying on timbre did not involve this area is not clear. In the study conducted by Scheich et al. (1998), both pitch and timbre were varied, whereas Brechmann et al. (2002) used frequency-modulated tones. Thus, it may be that TA activation only occurred because pitch-related processing was involved under competing stimulation conditions.

Due to the strong activation of T1 and T2 in the current pitch study compared to the timbre study, we were able to further subdivide these territories based on the foci of fMRI activation consistently found in all subjects. The more detailed spatial analysis revealed two clearly separable clusters of activation located in the lateral and medial part of each territory.

This further subdivision of the human auditory "core" and "lateral belt" region is in accordance with functional and anatomical data from non-human primates and humans. In the macaque monkey, a subdivision of the primary auditory cortex into two (A1, R) (Morel et al., 1993) or three (A1, R, RT) (Kaas and Hackett, 1998) fields is described functionally and anatomically. In humans, a subdivision of the primary auditory cortex along Heschl's gyrus into two or three separate fields was proposed by cytoarchitectonic studies (Hackett et al., 2001; Morosan et al., 2001; Wallace et al., 2002), intracortically recorded auditory event-related potentials

(Liegeois-Chauvel et al., 1994), and tonotopic gradients (Formisano et al., 2003). Thus, the subdivision of T1 into two or three areas is plausible.

Similarly, the lateral belt region in macaque monkeys can also be subdivided (CL, ML, AL, RTL) (Kaas and Hackett, 1998). In humans, cytoarchitectonic studies described two (Te 2.1, Te 2.2) (Morosan et al., 2005) or three subregions (PA, LA, STA) (Rivier and Clarke, 1997; Wallace et al., 2002) in this area, which postero-laterally adjoins the core area. Thus, the location of the subdivisions of T2 into a medial and lateral part is consistent with the cytoarchitectonic areas PA, LA, and Te 2.1, Te 2.2. A similar functional delineation of this region in Heschl's sulcus was shown by Hashimoto et al. (2000) using dichotic and binaural speech stimuli.

All subterritories of T1 and T2 showed a stronger BOLD response in the stream segregation condition compared to the control condition. However, the peaks in the averaged time course of BOLD responses were higher in the lateral subdivisions than in the medial ones (see Fig. 2D–G). This may be specific to pitch-related processing because numerous neuroimaging studies in humans (Griffiths et al., 1998; Hall et al., 2005; Patterson et al., 2002; Penagos et al., 2004; Schneider et al., 2005; Warren and Griffiths, 2003) have described the lateral part of Heschl's gyrus as sensitive to pitch processing. Similarly, in marmoset monkeys, pitch-selective neurons are found on the low-frequency borders of core areas A1 and R and belt areas AL and ML (Bendor and Wang, 2005). Using harmonic tones instead of pure tones, in streaming paradigms these neurons are likely candidates for exhibiting the kind of streaming-specific responses described for pure tones by Fishman et al. (2001, 2004). Within streaming paradigms, the modulation of the response of such neurons could be thought of in terms of top–down connections from higher order fields on the planum temporale. However, the functional role of these fields including the neural mechanisms related to stream segregation must be investigated in future electrophysiological studies on monkeys.

References

Baumgart, F., Kaulisch, T., Tempelmann, C., Gaschler-Markefski, B., Tegeler, C., Schindler, F., Stiller, D., Scheich, H., 1998. Electrodynamic headphones and woofers for application in magnetic resonance imaging scanners. Med. Phys. 25, 2068–2070.

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.

Behne, N., Scheich, H., Brechmann, A., 2005. Contralateral white noise selectively changes right human auditory cortex activity caused by a FM-direction task. J. Neurophysiol. 93, 414–423.

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

Bethmann, A., Tempelmann, C., De Bleser, R., Scheich, H., Brechmann, A., 2006. Determining language laterality by fMRI and dichotic listening. Brain Res. 1133, 145–157.

Binder, J.R., Frost, J.A., Hammeke, T.A., Rao, S.M., Cox, R.W., 1996. Function of the left planum temporale in auditory and linguistic processing. Brain 119, 1239–1247.

Binder, J.R., Liebenthal, E., Possing, E.T., Medler, D.A., Ward, B.D., 2004. Neural correlates of sensory and decision processes in auditory object identification. Nat. Neurosci. 7, 295–301.

Brechmann, A., Scheich, H., 2005. Hemispheric shifts of sound representation in auditory cortex with conceptual listening. Cereb. Cortex 15, 578–587.

Brechmann, A., Baumgart, F., Scheich, H., 2002. Sound-level-dependent representation of frequency modulations in human auditory cortex: a low-noise fMRI study. J. Neurophysiol. 87, 423–433.

Brechmann, A., Gaschler-Markefski, B., Sohr, M., Yoneda, K., Kaulisch, T., Scheich, H., 2007. Working memory specific activity in auditory cortex: potential correlates of sequential processing and maintenance. Cereb. Cortex 17, 2544–2552.

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

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

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

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

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

Efron, R., 1963a. The effect of handedness on the perception of simultaneity and temporal order. Brain 86, 261–284.

Efron, R., 1963b. Temporal perception, aphasia and D'ej'a Vu. Brain 86, 403–424.

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

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.

Formisano, E., Kim, D.S., Di Salle, F., van de Moortele, P.F., Ugurbil, K., Goebel, R., 2003. Mirror-symmetric tonotopic maps in human primary auditory cortex. Neuron 40, 859–869.

Gaschler, B., Schindler, F., Scheich, H., 1996. KHORFu: a KHOROS-based functional image post processing system. A statistical software package for functional magnetic resonance imaging and other neuroimage data sets. In: Prat, A. (Ed.), COMPSTAT 1996: Proceedings in Computational Statistics. 12th Symposium Held in Barcelona, Spain, 1996. Physica-Verlag, Heidelberg, pp. 57–58.

Griffiths, T.D., Warren, J.D., 2002. The planum temporale as a computational hub. Trends Neurosci. 25, 348–353.

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.

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., Oxenham, A.J., Micheyl, C., Wilson, E.C., Melcher, J.R., 2007. Human cortical activity during streaming without spectral cues suggests a general neural substrate for auditory stream segregation. J. Neurosci. 27, 13074–13081.

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.

Hall, D.A., Barrett, D.J., Akeroyd, M.A., Summerfield, A.Q., 2005. Cortical representations of temporal structure in sound. J. Neurophysiol. 94, 3181–3191.

Halperin, Y., Nachshon, I., Carmon, A., 1973. Shift of ear superiority in dichotic listening to temporally patterned nonverbal stimuli. J. Acoust. Soc. Am. 53, 46–50.

Hashimoto, R., Homae, F., Nakajima, K., Miyashita, Y., Sakai, K.L., 2000. Functional differentiation in the human auditory and language areas revealed by a dichotic listening task. Neuroimage 12, 147–158.

Kaas, J.H., Hackett, T.A., 1998. Subdivisions of auditory cortex and levels of processing in primates. Audiol. Neurootol. 3, 73–85.

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

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

Krumhansl, C.L., Iverson, P., 1992. Perceptual interactions between musical pitch and timbre. J. Exp. Psychol. Hum. Percept. Perform. 18, 739–751.

Leonard, C.M., Puranik, C., Kuldau, J.M., Lombardino, L.J., 1998. Normal variation in the frequency and location of human auditory cortex landmarks. Heschl's gyrus: where is it? Cereb. Cortex 8, 397–406.

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. Electroencephalogr. Clin. Neurophysiol. 92, 204–214.

Liegeois-Chauvel, C., de Graaf, J.B., Laguitton, V., Chauvel, P., 1999. Specialization of left auditory cortex for speech perception in man depends on temporal coding. Cereb. Cortex 9, 484–496.

Menon, V., Levitin, D.J., Smith, B.K., Lembke, A., Krasnow, B.D., Glazer, D., Glover, G.H., McAdams, S., 2002. Neural correlates of timbre change in harmonic sounds. Neuroimage 17, 1742–1754.

Micheyl, C., Carlyon, R.P., Cusack, R., Moore, B.C., 2004. Performance measure of auditory organization. In: Pressnitzer, D., de Cheveigne, A., McAdams, S., Collet, L. (Eds.), Auditory Signal Processing: Physiology, Psychoacoustics, and Models. Springer, New York, pp. 203–211.

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

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

Moore, B.C., Gockel, H., 2002. Factors influencing sequential stream segregation. Acta Acustica Acustica 88, 320–332.

Morel, A., Garraghty, P.E., Kaas, J.H., 1993. Tonotopic organization, architectonic fields, and connections of auditory cortex in macaque monkeys. J. Comp. Neurol. 335, 437–459.

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.

Morosan, P., Rademacher, J., Palomero-Gallagher, N., Zilles, K., 2005. Anatomical organization of the human auditory cortex: cytoarchitecture and transmitter receptors. In: König, R., Heil, P., Budinger, E., Scheich, H. (Eds.), The Auditory Cortex. A Synthesis of Human and Animal Research, vol. 3. LEA, Mahwah, NJ.

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.

Penhune, V.B., Zatorre, R.J., MacDonald, J.D., Evans, A.C., 1996. Interhemispheric anatomical differences in human primary auditory cortex: probabilistic mapping and volume measurement from magnetic resonance scans. Cereb. Cortex 6, 661–672.

Platel, H., Price, C., Baron, J.C., Wise, R., Lambert, J., Frackowiak, R.S., Lechevalier, B., Eustache, F., 1997. The structural components of music perception. A functional anatomical study. Brain 120, 229–243.

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

Rahne, T., Bockmann, M., von Specht, H., Sussman, E.S., 2007. Visual cues can modulate integration and segregation of objects in auditory scene analysis. Brain Res. 1144, 127–135.

Rahne, T., Deike, S., Selezneva, E., Brosch, M., Konig, R., Scheich, H., Bockmann, M., Brechmann, A., 2008. A multilevel and cross-modal approach towards neuronal mechanisms of auditory streaming. Brain Res. 1220, 118–131.

Rivier, F., Clarke, S., 1997. Cytochrome oxidase, acetylcholinesterase, and NADPH-diaphorase staining in human supratemporal and insular cortex: evidence for multiple auditory areas. Neuroimage 6, 288–304.

Scheich, H., Baumgart, F., Gaschler-Markefski, B., Tegeler, C., Tempelmann, C., Heinze, H.J., Schindler, F., Stiller, D., 1998. Functional magnetic resonance imaging of a human auditory cortex area involved in foreground–background decomposition. Eur. J. Neurosci. 10, 803–809.

Schneider, P., Sluming, V., Roberts, N., Scherg, M., Goebel, R., Specht, H.J., Dosch, H.G., Bleeck, S., Stippich, C., Rupp, A., 2005. Structural and functional asymmetry of lateral Heschl's gyrus reflects pitch perception preference. Nat. Neurosci. 8, 1241–1247.

Singh, P.G., 1987. Perceptual organization of complex-tone sequences: a tradeoff between pitch and timbre? J. Acoust. Soc. Am. 82, 886–899.

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

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., Steinschneider, M., 2006. Neurophysiological evidence for context-dependent encoding of sensory input in human auditory cortex. Brain Res. 1075, 165–174.

Sussman, E., Ritter, W., Vaughan Jr., H.G., 1998. Attention affects the organization of auditory input associated with the mismatch negativity system. Brain Res. 789, 130–138.

Swets, J.A., Tanner, W.P., Birdsall, T.G., 1961. Decision processes in perception. Psychol. Rev. 68, 301–340.

Tregellas, J.R., Davalos, D.B., Rojas, D.C., 2006. Effect of task difficulty on the functional anatomy of temporal processing. Neuroimage 32, 307–315.

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

Wallace, M.N., Johnston, P.W., Palmer, A.R., 2002. Histochemical identification of cortical areas in the auditory region of the human brain. Exp. Brain Res. 143, 499–508.

Warren, J.D., Griffiths, T.D., 2003. Distinct mechanisms for processing spatial sequences and pitch sequences in the human auditory brain. J. Neurosci. 23, 5799–5804.

Wetzel, W., Ohl, F.W., Scheich, H., 2008. Global versus local processing of frequency-modulated tones in gerbils: an animal model of lateralized auditory cortex functions. Proc. Natl. Acad. Sci. USA 105, 6753–6758.

Wilson, E.C., Melcher, J.R., Micheyl, C., Gutschalk, A., Oxenham, A.J., 2007. Cortical fMRI activation to sequences of tones alternating in frequency: relationship to perceived rate and streaming. J. Neurophysiol. 97, 2230–2238.

Woods, R.P., Grafton, S.T., Holmes, C.J., Cherry, S.R., Mazziotta, J.C., 1998. Automated image registration: I. General methods and intrasubject, intramodality validation. J. Comput. Assist. Tomogr. 22, 139–152.

Zatorre, R.J., Belin, P., Penhune, V.B., 2002. Structure and function of auditory cortex: music and speech. Trends Cogn. Sci. 6, 37–46.