Transient bold activity locked to perceptual reversals of auditory streaming in human auditory cortex and inferior colliculus

Stefan Schadwinkel and Alexander Gutschalk

J Neurophysiol 105:1977-1983, 2011. First published 16 February 2011; doi:10.1152/jn.00461.2010

J Neurophysiol 105: 1977–1983, 2011.
First published February 16, 2010; doi:10.1152/jn.00461.2010.

Transient bold activity locked to perceptual reversals of auditory streaming in human auditory cortex and inferior colliculus

Stefan Schadwinkel[1,2] and Alexander Gutschalk[1,2]
Departments of [1]Neurology and [2]Neuroradiology, University of Heidelberg, Heidelberg, Germany

Submitted 25 May 2010; accepted in final form 12 February 2011

Schadwinkel S, Gutschalk A. Transient bold activity locked to perceptual reversals of auditory streaming in human auditory cortex and inferior colliculus. *J Neurophysiol* 105: 1977–1983, 2011. First published February 16, 2010; doi:10.1152/jn.00461.2010.—Our auditory system separates and tracks temporally interleaved sound sources by organizing them into distinct auditory streams. This streaming phenomenon is partly determined by physical stimulus properties but additionally depends on the internal state of the listener. As a consequence, streaming perception is often bistable and reversals between one- and two-stream percepts may occur spontaneously or be induced by a change of the stimulus. Here, we used functional MRI to investigate perceptual reversals in streaming based on interaural time differences (ITD) that produce a lateralized stimulus perception. Listeners were continuously presented with two interleaved streams, which slowly moved apart and together again. This paradigm produced longer intervals between reversals than stationary bistable stimuli but preserved temporal independence between perceptual reversals and physical stimulus transitions. Results showed prominent transient activity synchronized with the perceptual reversals in and around the auditory cortex. Sustained activity in the auditory cortex was observed during intervals where the ΔITD could potentially produce streaming, similar to previous studies. A localizer-based analysis additionally revealed transient activity time locked to perceptual reversals in the inferior colliculus. These data suggest that neural activity associated with streaming reversals is not limited to the thalamo-cortical system but involves early binaural processing in the auditory midbrain, already.

auditory stream segregation; functional magnetic resonance imaging; interaural time differences; multistable perception

IN COMPLEX AUDITORY ENVIRONMENTS, like a noisy street or a crowded bar, multiple sources emit sound in parallel that reaches our ears all mixed together. It is on our brain to reconstruct the single sound sources from this mixture, a process referred to as "stream segregation" or "streaming" (Bregman 1990). Streaming can be studied with interleaved, repetitive tone patterns, which may either be integrated into a single auditory stream or perceptually segregated into two or more streams. This perceptual organization is to one part determined by several physical parameters such as sound repetition rate and differences between sound elements comprised in each sequence (Hartmann and Johnson 1991; Miller and Heise 1950; Moore and Gockel 2002). However, streaming perception is not completely determined by the physical stimulus but shows several characteristics of ambiguous, bistable perception (van Noorden 1975): while an integrated percept is generally experienced at the beginning of a stimulus sequence (Anstis and Saida 1985), streaming may build up if the physical separation between two sources lies beyond the "fission boundary" (van Noorden 1975), defined as the minimal (frequency) difference between sounds that is required to perceive streaming. Thereafter, depending on the degree of physical stimulus separation, streaming perception persists if the difference between streams is so large that an integration into a single stream is no longer possible. This threshold has been termed the "temporal coherence boundary" (van Noorden 1975). In the range between the fission and the temporal coherence boundaries, streaming is bistable and the perception may spontaneously switch back and forth between one- and two-stream perception after the initial buildup (Gutschalk et al. 2005; Pressnitzer and Hupé 2006).

Here, we adopted the original approach introduced by van Noorden (1975) to study bistable streaming reversals, in which the repetitive tone pattern is kept constant, while the stimulus parameters dissociating the individual tones are gradually changed over time. Interaural time differences (ΔITD) that produce perceptual lateralization were used as a streaming cue (Boehnke and Phillips 2005; Hartmann and Johnson 1991; Schadwinkel and Gutschalk 2010a, 2010b). In this paradigm, perceptual reversals towards stream segregation typically occur while the ΔITD is increased, whereas reversals back to an integrated percept are more likely to occur when the ΔITD is decreased again. Because the ΔITD used was mostly within the ambiguous range, the reversal time was not locked to any transitions of the physical stimulus. The advantage of this paradigm over one with a constant stimulus difference is that the time interval between reversals can be better controlled. In the present case, the inter-reversal time was set to a range that strongly reduced the overlap of hemodynamic responses evoked by subsequent perceptual reversals. BOLD functional (f)MRI activity was acquired within the auditory pathway, and the results showed transient activity phased locked to the perceptual reversals in the auditory cortex (AC) and in the inferior colliculus (IC), the auditory midbrain.

MATERIALS AND METHODS

Listeners. Twelve listeners (4 female; age 21–48 yr; mean 27) participated in fMRI measurements. None of the listeners had a history of central or peripheral hearing disorders. The study was performed in accordance with the Declaration of Helsinki, and the Ethics Committee of the Medical Faculty of Ruprecht-Karls-University Heidelberg approved the study.

Stimuli. The stimulus sequences comprised 125-ms harmonic tone complexes with an f_0 of 180 Hz (including 10-ms raised-cosine onset and offset ramps; sampling rate: 48 kHz; resolution: 16 bit; all 77 partials had the same amplitude and started in sine phase), digitally low-pass filtered at 5 kHz (6th-order Butterworth filter with zero phase shift). The complex tones were arranged temporally into a

Address for reprint requests and other correspondence: A. Gutschalk, Dept. of Neurology, Univ. of Heidelberg, Im Neuenheimer Feld 400, 69120 Heidelberg, Germany (e-mail: alexander.gutschalk@med.uni-heidelberg.de).

repeating ABBB pattern, where A and B represent tones with different ITD, according to their respective condition, with no silent gaps between consecutive tones.

The A tones were held at a constant lateralization to the left by an ITD of −687.5 μs. The B tones were lateralized to the right relative to the A tones, starting with a ΔITD of 41.76 μs, respective an absolute ITD of −645.83 μs. This ΔITD of 41.76 μs was used as a reference, at which the sequence was generally perceived as one integrated stream. Then, the ITD of the B tones was changed over time: the ITD was gradually increased for 32 s, held at a constant level for 2 s, and was gradually decreased back to the small reference ITD over 32 s. In one condition, the ITD of the B tones increased or decreased by one step (at 48 kHz, minimum possible ΔITD = 20.83 μs) every four quadruplets (INC4) or every six quadruplets (INC6). This yielded a maximum ITD of −312.5 μs (ΔITD = 375 μs) for INC4, and −416.67 μs (ΔITD = 270.83 μs) for INC6. Condition was varied pseudorandomly across sequences.

A localizer stimulus was used to map a region of interest (ROI) for IC activity. This stimulus comprised 32-s sequences of a repeating ABBB pattern with ITDs of −687.5 μs (A) and 0 μs (B), separated by a 32-s baseline without stimulation.

Procedures. Sound presentation was diotic via electrodynamic headphones (MR Confon, Magdeburg, Germany). Listeners were instructed to focus on the tones, to assume a ''one stream'' percept at the beginning of each run, and to indicate perceptual reversals as soon as possible. Subjects responded by pressing one of two keys on a hand response box (LUMItouch fMRI optical response keypad; Photon Control, Burnaby, BC, Canada), each button corresponding to one of the two percepts. Responses were recorded using Presentation (Neurobehavioral Systems, Albany, CA).

The sequences corresponding to the two variable-ITD conditions (INC4 and INC6) were presented in pseudorandom order during fMRI. The sequences were presented in four runs, each comprising eight 66-s sequences per run, yielding 16 presentations of each sequence. The 66-s sequences were separated by intervals in which the stimulation was continued with the constant reference ITD (ΔITD = 41.76 μs between A and B tones) for a pseudorandomized duration (24 to 40 s, step size: 1 s) to avoid subjects to anticipate a perceptual reversal based on the time interval. The 32-s localizer stimulus was presented five times with a 32-s baseline without stimulation in between. Time between runs was usually 30–60 s.

Scanner noise was attenuated by headphone earmuffs by ~45 dB and additional silicone earplugs (OHROPAX, Wehrheim, Germany), which provide attenuation of up to 23 dB. The overall sound pressure level of the stimulus was adjusted to be in the range of ~70 to 85 dB SPL at the ear.

Data acquisition. MRI data were acquired using a 3T scanner (Magnetom Trio; Siemens, Erlangen, Germany) equipped with an 8-channel phased-array head coil. Functional imaging was performed using a continuous echo-planar imaging sequence [gradient echo; repetition time, 1 s; echo time, 30 ms; flip angle, 90°; in-plane resolution, 64 × 64; field of view (FOV) 200 × 200 mm; 12 slices; slice thickness, 3.1 mm; gap, 1.023 mm]. The volume for functional imaging was chosen as 12 near-coronal slices, approximately perpendicular to the Sylvian fissure. This volume covered the AC from the posterior end of Planum temporale to the anterior aspect of the superior temporal gyrus, including the complete Heschl's gyrus in both hemispheres. Volume orientation was adjusted such that the medial geniculate bodies (MGB), the IC, and the brainstem were also covered.

Two structural magnetization prepared rapid gradient echo images of the whole head (sagittal in-plane resolution, 256 × 256; FOV, 256 × 256 mm; 192 slices; slice thickness, 1 mm) and a high-resolution T2-weighted structural image (in-plane resolution, 512 × 512; FOV, 200 × 200 mm; 12 slices; slice thickness, 3.1 mm) of the same volume as the functional image were acquired.

Data analysis. The fMRI data were motion corrected using AFNI (National Institute of Mental Health, Bethesda, MD; Cox and Jesmanowicz 1999). Correction of slice timing and coregistration was done using FSL (FMRIB, Oxford, UK; Jenkinson and Smith 2001; Smith 2002; Smith et al. 2004). Activity was detected by a general linear model using a single-gamma hemodynamic response function convolved with a boxcar function using the Functional Analysis Stream (FSFAST) from the FreeSurfer software package (Athinoula A. Martinos Center for Biomedical Imaging, Charlestown, MA). Drift components were modeled with second-order polynomial regressors. Motion correction parameters were included as regressors in the model. The remaining noise was modeled as a time-invariant linear AR(1) process.

Cortical activation was estimated by a random-effects multisubject analysis performed within the cortical surface space (Fischl et al. 1999), coregistered to the FreeSurfer brain template (fsaverage). For this purpose, the whole-head images were processed with Freesurfer to create an inflated projection of the cortical surface (Dale et al. 1999; Fischl et al. , 1999, 2001; Ségonne et al. 2004). The main analysis was restricted to the superior temporal plane, comprising mainly the AC. To correct for multiple comparisons, the positively activated vertices (determined by a vertex-wise uncorrected significance threshold) were clustered and significance levels of the clusters were estimated by Monte Carlo simulation of *z*-statistics (Ward 2000). The same *P* value was used for vertex-wise and cluster-wise thresholding in this process. Subcortical group activation was computed by a random-effects multisubject analysis in MNI-305 volume reference space and was not corrected for multiple comparisons.

The following contrasts were used: in the event-related analysis, ON was defined as 1-s event at the time of streaming onset and OFF was defined as 1-s event at the time of streaming offset, in correspondence with the listeners button-press responses. ON and OFF were contrasted against the whole stimulus run between the reversals (*ON* and *OFF* contrasts), comprising the constant reference interval as well as the variable-ITD intervals. Because an exploratory analysis could not identify differences between *ON* and *OFF* contrasts, the two were combined to a common *ON* + *OFF* contrast. In the blocked analysis (*PARADIGM*), the full 66-s interval of rising and falling ITD values was contrasted against the reference interval with static ΔITD = 41.76 μs.

Contrasts based on psychophysical response events were computed using an hemodynamic response function that peaked ~4 s after the response to account for an average response delay of ~2 s, i.e., parameters were as follows: $\delta = 1.5$; $\tau = 1.25$. All other contrasts were computed using standard parameters (Dale and Buckner 1997): $\delta = 2.25$; $\tau = 1.25$.

To extract activation time courses corresponding to the 66-s stimulus block, the respective group activation maps were projected back to the surface space of each subject. The reconstructed, intensity normalized, and averaged (across sequence repetitions per condition) BOLD-response time courses in the volume space were then projected into the surface space of the corresponding subject. From these projections, the time course of each ''active'' vertex was determined and averaged per subject. These per-subject time courses were then low-pass filtered (2nd-order Butterworth filter with zero phase shift) at half the Nyquist frequency (0.25 Hz), linearly detrended, and averaged to obtain an average group activation time course for each condition (INC4 and INC6) and hemisphere.

To reconstruct the activation time courses for the ON and OFF events, the time series for each voxel (within volume space) were at first low-pass interpolated (Institute of Electrical and Electronics Engineers 1979) to 10-ms resolution ($l = 4$, $\alpha = 0.5$, i.e., the input signal was assumed to be band limited with a cutoff frequency of half the Nyquist frequency). The time courses corresponding to the events were then averaged over a 40-s time window starting 16 s before the response (an 8-s preresponse baseline was subtracted from each epoch) and resampled to a 0.5-s resolution. The procedure was

performed separately for each condition (INC4 and INC6), event (ON and OFF), and subject. Time courses of subcortical ROIs were directly extracted from these data in the volume space; for cortical ROIs, the data were first projected into the cortical surface space and mapped to the group activation as above. Group activation time courses were thereby obtained for each condition, event, and hemisphere.

Average amplitudes and latencies of the BOLD transient following the ON and OFF events were determined by bootstrapping across subjects (Efron and Tibshirani 1997) with 10,000 resamples. For the cortical data, bootstrapping was performed for each event, each condition, and each ROI (averaged across hemispheres); for the subcortical data, bootstrapping was performed for each event (averaged across conditions and hemispheres to increase the signal-to-noise ratio). Bootstrap-*t*-intervals (Efron and Tibshirani 1997) were computed for all estimates.

An additional analysis of the BOLD waveforms was performed to estimate the degree to which sustained activity in the blocked *PARADIGM* analysis might potentially reflect summation of temporally spread perceptual reversals by modeling the time course during the 66-s stimulus block by convolving the time of the selected reversals (separately for each subject) with a convolution kernel (both sampled at a temporal resolution of 500 ms, i.e., one ABBB quadruplet) and dividing the result by the number of stimulus block repetitions. The convolution kernel was an idealized transient wave without negative undershoot and a peak latency of 4 s, which was matched to the results of the event-related analysis of reversals. Similar results were obtained when the original waves of the event-related analysis were used instead. ON and OFF events were modeled separately for each condition (INC4 and INC6) and for each ROI by scaling the convolution kernel by the bootstrap estimate of the respective average peak amplitude. The superposition of both modeled time courses was then used as the final model for the time course observed during the stimulus block.

To estimate if the waveforms comprised additional sustained activity not explained by this model, the modeled time course was subtracted from its corresponding measured time course. This analysis was performed in the interval between 24 and 42 s after stimulus-block onset, where the ΔITD was maximal. In this interval, streaming was expected most frequently, whereas the majority of perceptual reversals happened before and after this interval. The average amplitude in this interval was estimated by bootstrapping across subjects with 10,000 resamples, before and after subtracting the model. Bootstrap-*t*-intervals were computed to test whether the average amplitude was different from zero. In addition, difference waveforms (measured minus modeled time courses) were calculated and *t*-intervals were determined on a per-sample basis by bootstrapping with 10,000 resamples.

RESULTS

Behavioral data. Listeners indicated streaming onset (one- to two-stream reversal, ON) and offset (two- to one-stream reversal, OFF) with two different keys, while they listened in the scanner to the constantly repeating ABBBABBB… pattern, where A was constantly lateralized to the left (ITD = −687.5 μs) and B slowly changed lateralization in the course of 66 s from −645.83 μs to −312.5 μs (INC4) or to −416.67 μs (INC6). Figure 1 shows the distribution of the listeners responses indicating streaming onset and offset. The distribution of reversal responses show that, while the ON responses were mostly constrained to the rising and the OFF responses to the falling ITD slopes, they occurred in a widely spread time interval distributed over >10 s (mean of subject-level interquartile ranges: INC4:ON = 13.9 s; INC6:ON = 12.4 s; INC4:OFF = 8.6 s; INC6:OFF = 8.3 s). This broad distribution clearly demonstrates that the occurrence of perceptual reversals was constrained but not determined by the physical stimulus.

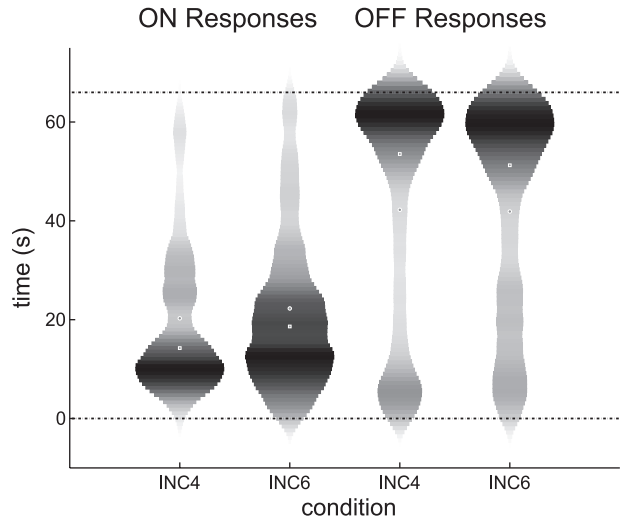


Fig. 1. Distribution plots of the response times pooled over all subjects (n = 12), separately for onset (ON, *left*) and offset (OFF, *right*) of perceiving segregated streams. Response times are plotted relative to the onset of the 66-s stimulus block. Circles indicate the mean; squares indicate the median. Intensity and widths consecutively represent the relative number of respective responses for each instance in time. After considering a response delay of ~2 s, the actual change in interaural time difference (ΔITD) corresponding to the highest frequency of reversals would be estimated around 100 μs for both directions (ON and OFF). INC4, ITD increase/decrease every 4 quadruplets; INC6, every 6 quadruplets.

Cortical activation detected by fMRI contrast analysis. The analysis of cortical activity was performed on the cortical surface and restricted to the superior temporal plane (Fig. 2), because the volume for functional imaging was restricted to cover the AC and the brainstem. Two types of analysis were performed on the data: sustained activity was detected by a contrast of the 66-s stimulus block in which streaming occurred against the reference interval (*PARADIGM*; Fig. 2, *left*). This contrast revealed activity in most parts of the AC, including the posterior Heschl's gyrus (HG), the planum temporale parts of the superior temporal gyrus, and small areas anterior to HG. A second analysis (*ON + OFF*; Fig. 2, *right*) was used to explore activity related to the perceptual reversals, based on the subjects' responses summarized in Fig. 1. This analysis contrasted the ON and OFF events against the whole stimulus run. The activity revealed by *ON + OFF* covered the same area as the *PARADIGM* contrast but extended more into the planum polare and towards the insula and lower parietal lobe. Additionally, there was activity in the superior temporal sulcus on the right hemisphere.

Cortical BOLD response time courses. The difference in the distribution of activity described in the previous section suggests different temporal activation patterns, i.e., a sustained activation pattern during increased ΔITD intervals that is limited to the central AC and a transient activation pattern, temporally related to perceptual reversals, that is present in the central AC but also in surrounding areas. To access these different activation patterns in more detail, the activation cluster was separated into a central and an outer ROI, and the corresponding fMRI time courses were determined. These two

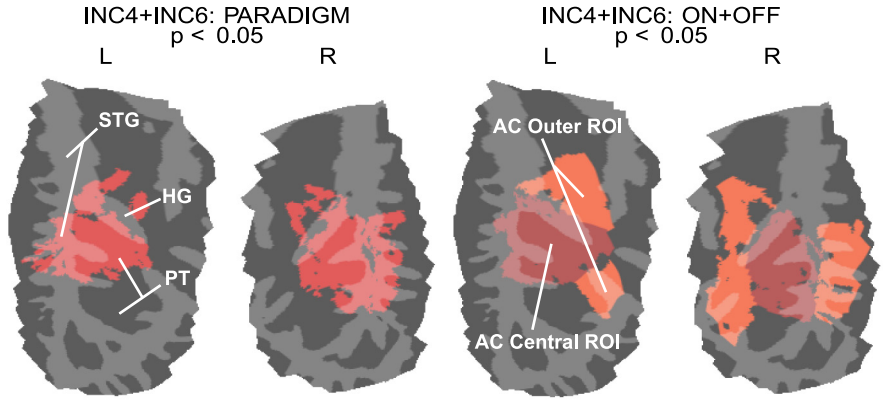


Fig. 2. Significant AC activation detected for the *PARADIGM* (*left*) and the *ON + OFF* (*right*) contrasts ($n = 12$; cutoff $P < 0.05$), mapped on the flattened cortical surface of the superior temporal plane. All maps depict the result of a random-effects group analysis, corrected for multiple comparisons by clustering positively activated vertices based on Monte Carlo simulation of *z*-statistics (Ward 2000). Orange and red in the ON + OFF map (*right*) denote which vertices were included in the central (red) or the outer (orange) auditory cortex (AC) region of interest (ROI). Both areas (red and orange) were detected together as one single cluster with the same statistical threshold, and their separation was later on defined according to anatomical boundaries. Locations of Heschl's gyrus (HG), the Planum temporale (PT), and the superior temporal gyrus (STG) are indicated on the *leftmost* patch.

ROIs are depicted in Fig. 2, *right*, as red (inner) and orange (outer) areas. Note, that both areas were equally significant as one single cluster and the separation was introduced afterwards on an anatomical basis for further analysis. The time courses of the BOLD response in the two AC ROIs, based on the *ON + OFF* contrast, are depicted in Fig. 3: over the 66-s stimulus block, the activity in the central ROI exhibited distinct sustained activity, while the activity level in the outer ROI was comparatively low. Transient activity, time-locked to the ON and OFF events, showed peaks ~4 s that were equally pronounced in both ROIs. There was no significant difference between ON and OFF (based on two-tailed bootstrap-*t*-intervals at $P < 0.05$; see Fig. 3, *bottom row*).

To estimate the degree to which the transient activity associated with reversals may account for parts of the observed sustained signal throughout the stimulus block, reversals indicated by the listeners were convolved with a transient response matched to the one observed in the event-related analysis. The modeled time courses were then subtracted from the measured time courses (separately for each subject), and the average sustained amplitude (measured between 24 and 42 s after stimulus onset) was determined by bootstrapping (see Fig. 3, *bottom*). The results showed that there was significant sustained activity not explained by the model in the central ROI but not in the outer ROI (Fig. 3).

Subcortical activity. For the evaluation of subcortical activity in the auditory pathway, the analysis was performed in a volume-based-reference space. Because the perceptual reversals were not temporally locked to the physical stimulus, they can only be studied when their occurrence is indicated by the listener. Thus we expect that the *ON + OFF* contrast would not only reveal activity related to the perceptual reversals in the auditory pathway but also activity related to the motor task. The analysis presented here was therefore restricted to the canonical auditory pathway, where the influence of motor and task-related activity is thought to be less likely than in other cortical and brainstem centers. While the event-related contrast of the reversals (*ON + OFF*) revealed widespread activity in the thalamus and midbrain, this activity was clustered together and clearly included sites outside of the auditory pathway. Therefore, the subcortical activity revealed by the *ON + OFF* contrast was not analyzed in more detail.

To separate auditory centers more specifically, a functional ROI based on a localizer stimulus was used instead, based on a 32-s ABBB pattern contrasted vs. silence and presented to all subjects. The contrast of the 32-s localizer-stimulus block against a silent baseline (*LOCALIZER*) reliably detected activity in the IC at $P < 0.01$ (not corrected for multiple comparisons) and the AC but not in the MGB. To confirm that activity was located in the IC, the group activity was projected back on the individual T2-weighted high resolution images. The analysis confirmed that the functional ROI accurately matched the anatomical location of the IC in all subjects.

The ROI based on the functional-localizer activity (Fig. 4, *top row*) was then used to extract time courses for the ON and OFF events in the IC. Data were averaged across hemispheres and gradient conditions to improve the signal-to-noise ratio. The resulting time courses are depicted in Fig. 4, *bottom row*.

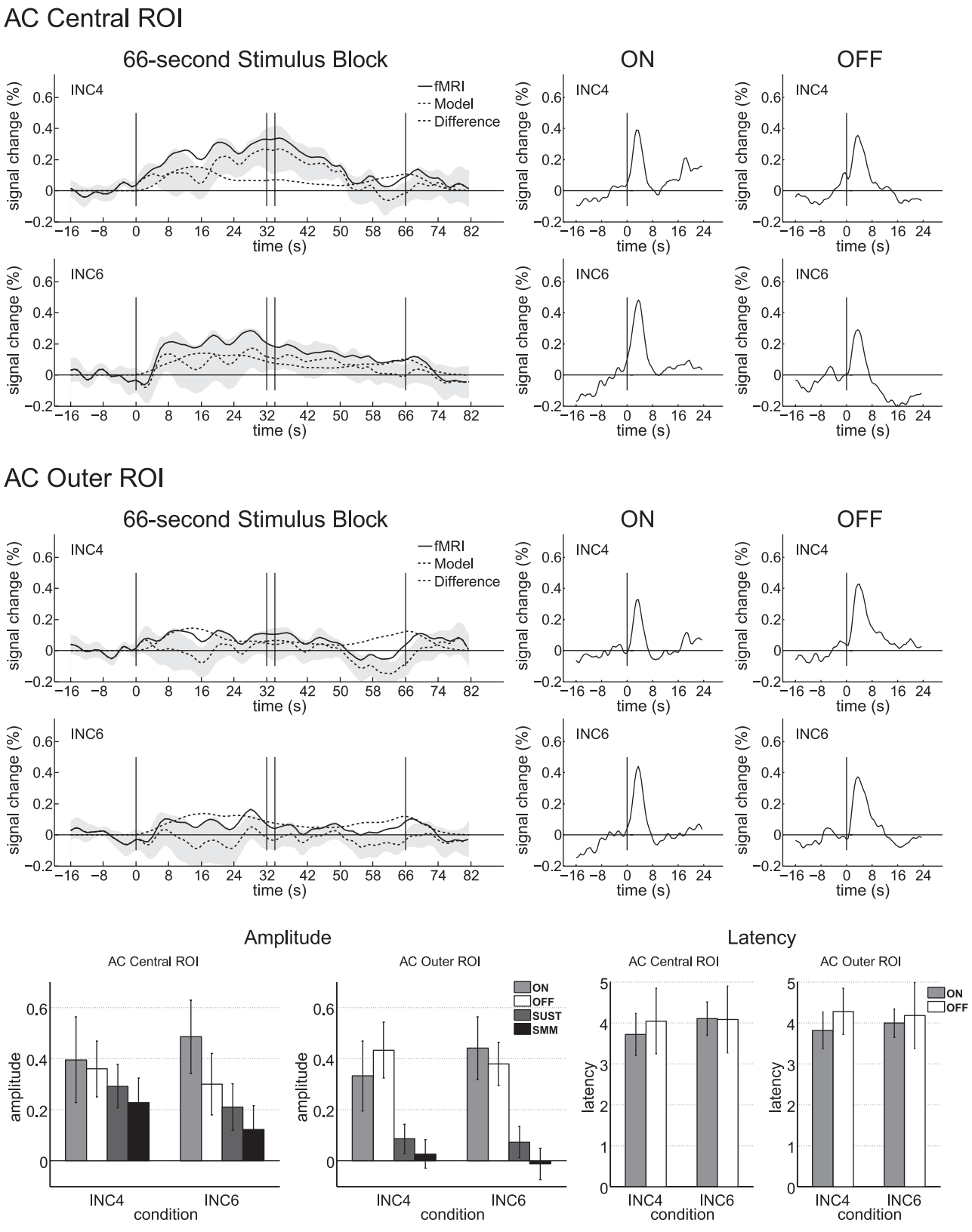
Despite the higher noise level, the time course corresponding to the ON event was similar to the transients in the AC and was significant with respect to bootstrap-based *t*-intervals [peak amplitude (%signal change) ± bootstrap-*t*-interval for $P < 0.05$, two-tailed: 0.39 ± 0.15]. The response to the OFF event was not

Fig. 3. Time courses with corresponding bootstrap-based amplitude and latency estimates of the functional (f)MRI activation averaged across subjects ($n = 12$). Data for both gradients (INC4 and INC6) are displayed separately. *Top*: fMRI activation time courses in the central AC ROI. *Middle*: time courses in the outer AC ROI. *Leftmost* time courses depict a linearly detrended average over the 66-s stimulus block (*fMRI*), a time course that was computationally modeled based on psychophysical responses (*Model*; see METHODS for details), and a per-sample bootstrap-based difference between the *fMRI* and *Model* time courses (*Difference*) with per-sample bootstrap-*t*-intervals ($P < 0.05$) in grey shading. Two time courses to the *right* are event-related averages corresponding to the ON and OFF events (without detrending). *Bottom*: bootstrap estimates of the amplitudes of the transient peaks (ON, OFF), the sustained activity measured between 24 and 42 s during the stimulus block (SUST), and the difference between measured sustained activity and its respective model (SUST MINUS MODEL, SMM) in the central and outer ROI. Latencies of the BOLD transients are shown on the *bottom right*. All bars denote two-sided bootstrap-*t*-intervals at $P < 0.05$.

as distinctive but also showed a significant peak (0.36 ± 0.14). The average peak latencies were 3.45 s (ON) and 3.7 s (OFF), but there was no significant latency difference between ON and OFF in IC nor between the IC and the cortical ROIs (see Fig. 3) based on two-tailed bootstrap-*t*-intervals at $P < 0.05$.

DISCUSSION

The present study introduces a modified streaming paradigm for fMRI in which one source is slowly moved away from and back to the position of a second, static source, which emits a temporally interleaved tone stream. This configuration is usu-



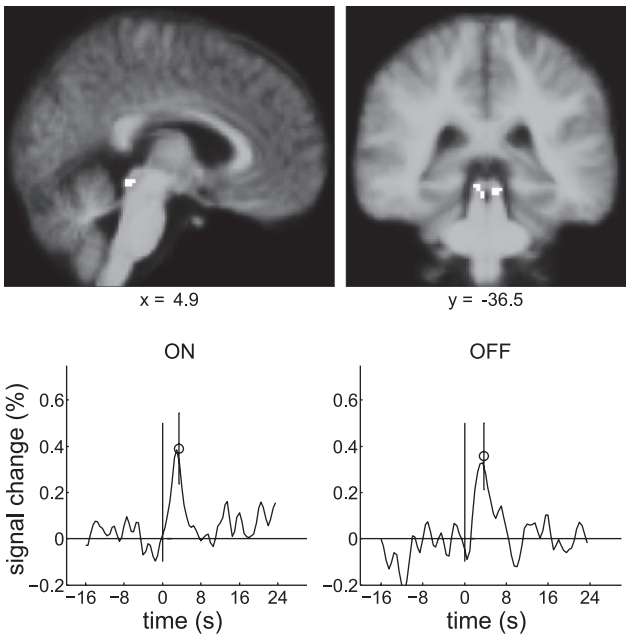


Fig. 4. Significant activation (depicted in white) in the inferior colliculus (IC) as detected by the *LOCALIZER* contrast, cutoff $P < 0.01$, not corrected for multiple comparisons. All activation outside IC was masked. Talairach coordinates of the centroids of the activity were at (x; y; z): *left*: (−5.5; −37.7; −6.7), *right*: (4.6; −35.2; −6.4); Talairach coordinates of the depicted slices are given below the respective image. *Bottom*: fMRI activation time courses corresponding to the ON and OFF events in the IC, extracted from the voxels detected by the *LOCALIZER* contrast (as shown at *top*). Time courses are averages across hemispheres, gradient conditions, and subjects ($n = 12$). Bootstrap estimates of the average peak amplitude and latency of the respective transient are depicted by a grey circle; bars indicate two-sided bootstrap-t-intervals at $P < 0.05$.

ally perceived as a single tone stream when the two sources are close to each other and as two segregated streams when the spatial separation is more distinct. Because of the slow change of lateralization, the reversals between the two perceptual organizations (one or two streams) were constrained by the direction of the movement, but not tightly locked to a specific lateralization difference. This setup allowed for a separation of activity associated with perceptual reversals, similar to bistable paradigms using static stimulus configurations (Gutschalk et al. 2005; Kondo and Kashino 2009; Pressnitzer and Hupé 2006). In contrast to static bistable streaming paradigms, however, the slowly changing stimulus allowed us to better control the time interval between subsequent reversals (see Fig. 1). This is of interest, because spontaneous reversals occur on average after 4–8 s (Gutschalk et al. 2005; Kondo and Kashino 2009; Pressnitzer and Hupé 2006), which is comparatively short and produces considerable overlap of the BOLD activity associated with subsequent reversals. The longer interval (on average 54.4 s) between subsequent reversals in the present study allowed us to reconstruct the time courses of the activity related to the perceptual reversals without the need for prior modeling assumptions.

Our results reveal transient BOLD activity in the AC and midbrain that was locked to the reversal time but not to the physical stimulus, and we found no significant differences between the direction of perceived reversals. The cortical part of this activity is similar to the activity observed by Kondo and Kashino (2009), who used spontaneous reversals in a static streaming paradigm with tone-frequency differences. These authors suggested that the activity was related to the detection of a perceptual change, because they found similar activation in the AC and thalamus when their listeners detected a physically defined target stimulus in an oddball paradigm. While we also found thalamic activity in our study, we were not able to accurately separate activity within the MGB from activity in the neighboring thalamic nuclei and therefore omitted the analysis of this stage. However, we found additional transient BOLD activity at an earlier stage of the auditory pathway in the IC, which was successfully separated by a functional localizer.

While the transient activity associated with streaming reversals is related to a perceptual rather than a physical event, the additionally observed sustained BOLD activity in AC appears to be more tightly coupled to the separation of the sound sources in the stimulus. Sustained activity in the AC is determined by tone duration, repetition rate, and interstimulus interval (Giraud et al. 2000; Harms et al. 2005; Harms and Melcher 2002) and may be related to feature-selective adaptation, which has been suggested to be closely related to streaming perception (Fishman et al. 2001; Micheyl et al. 2005; Pressnitzer et al. 2008). Accordingly, human fMRI studies found streaming-associated sustained signal enhancement in instances where the perceived tone-repetition rate decreased due to stream segregation (Gutschalk et al. 2007; Schadwinkel and Gutschalk 2010; Wilson et al. 2007). Covariation between streaming perception and ongoing activity in the AC beyond selective adaptation has been demonstrated with MEG (Gutschalk et al. 2005), but the size of this effect was considerably smaller than selective adaptation. Using fMRI, Cusack (2005) found enhanced activity for bistable streaming in the intraparietal sulcus but not in the AC.

Potential underpinnings of reversal associated BOLD transients. The nature of the transient BOLD activity in the IC as well as the AC is currently unclear. Since both structures represent predominantly sensory areas, a bottom-up mechanism that is directly related to the mechanisms involved in streaming representation was generally plausible: Pressnitzer et al. (2008) demonstrated that recordings from the cochlear nucleus in guinea pigs can predict behavioral measures of streaming in humans, suggesting that subcortical structures may already be involved in auditory scene analysis. A role for the cochlear nucleus, where processing is mostly monaural, is unlikely to contribute to streaming based on ITDs, however, which may be first processed in the medial superior olive (Palmer 2004). The IC is a subsequent, obligatory center in the ascending auditory system, and there is plenty of evidence of its importance for ITD representations (Hancock and Delgutte 2004; Thompson et al. 2006). As such, the IC may well be involved in early bottom-up processing of stream segregation based on ITDs and other spatial cues. This would be similar to the visual system, where subcortical representations of bistable perception during binocular rivalry have been demonstrated as early as the lateral geniculate nucleus (Haynes et al. 2005; Wunderlich et al. 2005), which is the first visual processing stage subsequent to the retina.

On the other hand, top-down projections from the AC control IC activity based on selective attention (Rinne et al. 2008; Suga and Ma 2003), modify tuning for space and frequency (Gao and Suga 1998; Nakamoto et al. 2008), and disrupt learning-induced plasticity if lesioned (Bajo et al. 2009). Therefore, the transient activity that co-occurs with streaming reversals may alternatively reflect top-down processing, for example, related to attentional orienting associated

with the detection of a second sound source. Such a process could hypothetically initiate a cortically controlled adaptation of IC neurons' receptive fields to enhance the separation of auditory streams into distinct neuronal populations (Fishman et al. 2001; Micheyl et al. 2005; Pressnitzer et al. 2008).

In conclusion, these data demonstrate IC and AC activity that is not determined by the physical stimulus but is associated with a perceptual or cognitive event related to streaming reversals. Especially the presence of such activity in the IC is of great interest for the refinement of neural models of auditory stream segregation. It will therefore be important to understand if the reversal-related activity is generated by top-down or bottom-up processes.

GRANTS

This research was supported primarily by Deutsche Forschungsgemeinschaft Grant GU593/3–1 and additionally by Bundesministerium für Bildung und Forschung Grant 01EV0712.

DISCLOSURES

No conflicts of interest, financial or otherwise, are declared by the author(s).

REFERENCES

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

Bajo VM, Nodal FR, Moore DR, King AJ. The descending corticocollicular pathway mediates learning-induced auditory plasticity. *Nat Neurosci* 13: 253–260, 2009.

Boehnke SE, Phillips DP. The relation between auditory temporal interval processing and sequential stream segregation examined with stimulus laterality differences. *Percept Psychophys* 67: 1088–1101, 2005.

Bregman AS. *Auditory Scene Analysis*. Cambridge: MIT Press, 1990.

Cox RW, Jesmanowicz A. Real-time 3d image registration for functional MRI. *Magn Reson Med* 42: 1014–1018, 1999.

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

Dale AM, Buckner RL. Selective averaging of rapidly presented individual trials using fMRI. *Hum Brain Mapp* 5: 329–340, 1997.

Dale AM, Fischl B, Sereno MI. Cortical surface-based analysis. I. Segmentation and surface reconstruction. *Neuroimage* 9: 179–194, 1999.

Efron B, Tibshirani RJ. *An Introduction to the Bootstrap*. New York: Chapman & Hall, 1997.

Fischl B, Liu A, Dale AM, Center NMR, Hosp MG, Boston MA. Automated manifold surgery: constructing geometrically accurate and topologically correct models of the human cerebral cortex. *IEEE Trans Med Imaging* 20: 70–80, 2001.

Fischl B, Sereno MI, Dale AM. Cortical surface-based analysis II: inflation, flattening, and a surface-based coordinate system. *Neuroimage* 9: 195–207, 1999.

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

Gao E, Suga N. Experience-dependent corticofugal adjustment of midbrain frequency map in bat auditory system. *Proc Natl Acad Sci USA* 95: 12663–12670, 1998.

Giraud AL, Lorenzi C, Ashburner J, Wable J, Johnsrude I, Frackowiak R, Kleinschmidt A. Representation of the temporal envelope of sounds in the human brain. *J Neurophysiol* 84: 1588–1598, 2000.

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

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

Hancock KE, Delgutte B. A physiologically based model of interaural time difference discrimination. *J Neurosci* 24: 7110–7117, 2004.

Harms MP, Guinan JJ, Sigalovsky IS, Melcher JR. 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, 2005.

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

Hartmann WM, Johnson D. Stream segregation and peripheral channeling. *Music Perception* 9: 155–184, 1991.

Haynes JD, Deichmann R, Rees G. Eye-specific effects of binocular rivalry in the human lateral geniculate nucleus. *Nature* 438: 496–499, 2005.

Institute of Electrical and Electronics Engineers. *Programs for Digital Signal Processing*. New York: IEEE Press, 1979.

Jenkinson M, Smith S. A global optimisation method for robust affine registration of brain images. *Med Image Anal* 5: 143–156, 2001.

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

Micheyl C, Tian B, Carlyon RP, Rauschecker JP. Perceptual organization of tone sequences in the auditory cortex of awake macaques. *Neuron* 48: 139–148, 2005.

Miller GA, Heise GA. The trill threshold. *J Acoust Soc Am* 22: 637–638, 1950.

Moore BCJ, Gockel H. Factors influencing sequential stream segregation. *Acta Acust United Acust* 88: 320–333, 2002.

Nakamoto KT, Jones SJ, Palmer AR. Descending projections from auditory cortex modulate sensitivity in the midbrain to cues for spatial position. *J Neurophysiol* 99: 2347–2356, 2008.

Palmer AR. Reassessing mechanisms of low-frequency sound localisation. *Curr Opin Neurobiol* 14: 457–460, 2004.

Pressnitzer D, Hupé JM. Temporal dynamics of auditory and visual bistability reveal common principles of perceptual organization. *Curr Biol* 16: 1351–1357, 2006.

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

Rinne T, Balk MH, Koistinen S, Autti T, Alho K, Sams M. Auditory selective attention modulates activation of human inferior colliculus. *J Neurophysiol* 100: 3323–3327, 2008.

Ségonne F, Dale AM, Busa E, Glessner M, Salat D, Hahn HK, Fischl B. A hybrid approach to the skull stripping problem in MRI. *Neuroimage* 22: 1060–1075, 2004.

Schadwinkel S, Gutschalk A. Activity associated with stream segregation in human auditory cortex is similar for spatial and pitch cues. *Cereb Cortex* 20: 2863–2873, 2010a.

Schadwinkel S, Gutschalk A. Functional dissociation of transient and sustained fMRI BOLD components in human auditory cortex revealed with a streaming paradigm based on interaural time differences. *Eur J Neurosci* 32: 1970–1978, 2010b.

Smith SM. Fast robust automated brain extraction. *Hum Brain Mapp* 17: 143–155, 2002.

Smith SM, Jenkinson M, Woolrich MW, Beckmann CF, Behrens TEJ, Johansen-Berg H, Bannister PR, Luca MD, Drobnjak I, Flitney DE, Niazy RK, Saunders J, Vickers J, Zhang Y, Stefano ND, Brady JM, Matthews PM. Advances in functional and structural MR image analysis and implementation as FSL. *Neuroimage* 23: 208–219, 2004.

Suga N, Ma X. Multiparametric corticofugal modulation and plasticity in the auditory system. *Nat Rev Neurosci* 4: 783–794, 2003.

Thompson SK, von Kriegstein K, Deane-Pratt A, Marquardt T, Deichmann R, Griffiths TD, McAlpine D. Representation of interaural time delay in the human auditory midbrain. *Nat Neurosci* 9: 1096–1098, 2006.

van Noorden LPAS. *Temporal Coherence in the Perception of Tone Sequences.* Eindhoven, The Netherlands: Eindhoven University of Technology, 1975.

Ward BD. Simultaneous inference for fMRI data. *AFNI Documentation*. Milwaukee, WI: Medical College of Wisconsin, 2000.

Wilson EC, Melcher JR, Micheyl C, Gutschalk A, Oxenham AJ. Cortical fMRI activation to sequences of tones alternating in frequency: relationship to perceived rate and streaming. *J Neurophysiol* 97: 2230, 2007.

Wunderlich K, Schneider KA, Kastner S. Neural correlates of binocular rivalry in the human lateral geniculate nucleus. *Nat Neurosci* 8: 1595–1602, 2005.