

# Involvement of the Thalamocortical Loop in the Spontaneous Switching of Percepts in Auditory Streaming

Hirohito M. Kondo<sup>1</sup> and Makio Kashino<sup>1,2,3</sup>

<sup>1</sup>NTT Communication Science Laboratories, NTT Corporation, and <sup>2</sup>Exploratory Research for Advanced Technology Shimojo Implicit Brain Function Project, Japan Science and Technology Agency, Atsugi, Kanagawa 243-0198, Japan, and <sup>3</sup>Interdisciplinary Graduate School of Science and Engineering, Tokyo Institute of Technology, Yokohama, Kanagawa 226-8503, Japan

Perceptual grouping of successive frequency components, namely, auditory streaming, is essential for auditory scene analysis. Prolonged listening to an unchanging triplet-tone sequence produces a series of illusory switches between a single coherent stream (S1) and two distinct streams (S2). The predominant percept depends on the frequency difference ( $\Delta f$ ) between high and low tones. Here, we combined the use of different  $\Delta f$ s with an event-related fMRI design to identify whether the temporal dynamics of brain activity differs depending on the direction of perceptual switches. The results demonstrated that the activity of the medial geniculate body (MGB) in the thalamus occurred earlier during switching from nonpredominant to predominant percepts, whereas that of the auditory cortex (AC) occurred earlier during switching from predominant to nonpredominant percepts, regardless of  $\Delta f$ . The asymmetry of temporal precedence indicates that the MGB and AC activations play different roles in perceptual switching and depend on perceptual predominance rather than on S1 and S2 percepts per se. Our results suggest that feedforward and feedback processes in the thalamocortical loop are involved in the formation of percepts in auditory streaming.

## Introduction

Humans and animals possess the sophisticated ability to integrate complex acoustic inputs into one organized percept and switch their attention among several auditory streams within inputs. For instance, we can hear a sound mixture as music and discern the sounds of individual instruments of an orchestra. The sequential integration and segregation of frequency components for the formation of percepts, which is called auditory streaming, is essential for auditory scene analysis. The psychophysical factors influencing auditory streaming are known (van Noorden, 1975; Moore and Gockel, 2002). However, the neural correlates of auditory streaming remain unclear.

Several studies have tried to clarify the neural correlates of auditory streaming (Alain et al., 1998; Sussman et al., 1999; Fishman et al., 2001, 2004; Näätänen et al., 2001). However, a common limitation of these studies is that neural correlates corresponding to the different percepts were evoked by physically different stimuli. This makes it difficult to determine whether the neural correlates reflect auditory percepts or the physical properties of the stimuli. To overcome this difficulty, recent studies have used the build-up of auditory streaming, where a physical unchanging sequence of triplet tones initially tends to be heard as a single coherent stream (S1) and is split into two distinct streams (S2) after several seconds. Neuro-

physiological studies have shown that patterns of neuronal responses in the primary auditory area of awake rhesus monkeys and in the cochlear nucleus of anesthetized guinea pigs are correlated with changes in predominant percepts from S1 to S2 in humans (Micheyl et al., 2005; Pressnitzer et al., 2008). An event-related potential (ERP) study has revealed that responses at the midline frontocentral (FCz) area in humans are correlated with percepts during the build-up (Snyder et al., 2006). A functional magnetic resonance imaging (fMRI) study has demonstrated that the intraparietal sulcus is more highly activated during S2 than during S1 (Cusack, 2005), whereas a magnetoencephalography (MEG) study has indicated that the amplitude of  $N_{1m}$  response in the nonprimary auditory areas reflects frequency difference ( $\Delta f$ ) between high (H) and low (L) tones (Gutschalk et al., 2005). However, although these studies have clarified brain activity related to the “representation” of percepts in auditory streaming, the findings regarding activation areas are rather divergent, pointing to the need for some other line of research.

The present study focused on the “switching” of percepts after the build-up to gain new insights into the issue of where and how auditory percepts are formed in the brain. Although perceptual predominance after the build-up depends on  $\Delta f$ , its effects on brain activity are not clear. We used two  $\Delta f$  conditions with different perceptual predominance and examined whether brain activity differs between the two directions of perceptual switches. We hypothesized that if some brain area is causally involved in perceptual switching, then the response in that area should precede that in other areas. We further examined whether temporal precedence of activations is affected by perceptual predominance or auditory percepts (i.e., S1 and S2).

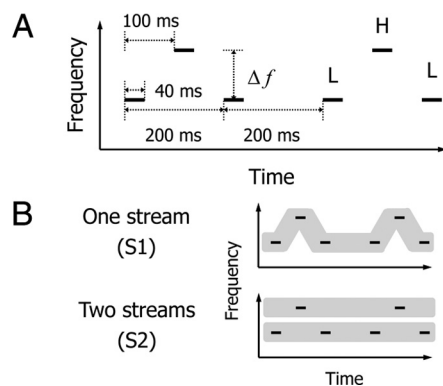
Received April 1, 2009; revised Aug. 11, 2009; accepted Aug. 29, 2009.

We thank Daniel Pressnitzer and Christophe Micheyl for their thoughtful comments on an earlier version of this manuscript.

Correspondence should be addressed to Hirohito M. Kondo, NTT Communication Science Laboratories, NTT Corporation, 3-1 Morinosato Wakamiya, Atsugi, Kanagawa 243-0198, Japan. E-mail: hkondo@brl.ntt.co.jp.

DOI:10.1523/JNEUROSCI.1549-09.2009

Copyright © 2009 Society for Neuroscience 0270-6474/09/2912695-07\$15.00/0



**Figure 1.** Schematic representation of stimuli and two types of auditory percepts. **A**, A sound sequence (a total of 90 s) of a triplet tone was presented in the perceptual switching task. Participants performed this task at either  $\Delta f \approx 2$  or 6 semitones. H and L stand for high- and low-frequency tones. **B**, Participants pressed a button with their left thumb when they perceived subjective changes from one stream (S1, LHL-LHL-...) to two streams (S2, L-L-... and H-H-...) or those from S2 to S1.

## Materials and Methods

**Participants.** Thirty-two people participated in a psychophysical experiment. They were right-handed Japanese college students with normal hearing. The mean score of the Edinburgh Handedness Inventory was 92.9 (range 80–100). None had any history of neurological and psychiatric illness. On the basis of the results of the psychophysical experiment, we recruited 24 participants (12 males and 12 females; mean age 23.4 years, range 19–30 years) for an fMRI experiment. All participants gave written informed consent, which was approved by the Ethics and Safety Committees of NTT Communication Science Laboratories and ATR Institute International.

**Stimuli and task procedures.** The psychophysical experiment comprised five 90 s runs of the perceptual switching task at  $\Delta f \approx 2$  semitones or the same ones at  $\Delta f \approx 6$  semitones. In the perceptual switching task, stimuli were 225 repetitions of a triplet tone that comprised L and H tones with intervals of silence (Fig. 1A). The  $\Delta f$ s between L and H tones were  $\sim 2$  and 6 semitones, while the center frequency was fixed to 1 kHz. To avoid harmonic consonance between L and H tones, the frequency of each tone was set to the following values: L = 937 Hz and H = 1069 Hz at  $\Delta f \approx 2$  semitones; L = 823 Hz and H = 1213 Hz at  $\Delta f \approx 6$  semitones. The duration of each tone was 40 ms, which included rising and falling cosine ramps of 10 ms. Participants were asked to listen to the tone sequence passively without any particular focus and attitude. They pressed a button with their left thumb to indicate that they had perceived subjective changes between S1 with a galloping rhythm and S2 with an isochronous rhythm for each stream (Fig. 1B). The stimuli were presented at 65 dB SPL diotically through headphones.

The fMRI experiment consisted of alternation of 90 s runs between the perceptual switching and tone detection tasks. We assumed the tone detection task to be an oddball task, which is to detect rare sounds embedded in a sequence of standard sounds. The stimuli in the tone detection task were the same as those in the perceptual switching task except that a tone pip was emitted using the identical temporal sequence of perceptual switching reported in the previous run. The tone pip at 1 kHz with onset and offset ramps of the cosine curve had the total duration of 20 ms. Participants were instructed to press a button when the tone pip was presented on the background of the triplet-tone sequence so that the number of responses for detection of tone pips would match as closely as possible that for the detection of perceptual switching. Thus, the detection of perceptual changes could be dissociated from that of physical changes in constant stimulation. We used the software Presentation (Neurobehavioral Systems) to synchronize the presentation timing of the stimuli with the scanner sequence.

**Data acquisition.** Images were acquired with a 1.5 T MRI scanner (Magnex Eclipse; Shimadzu-Marconi) with a standard head coil. Head

motion was minimized with a forehead strap and comfortable padding around the participant's head. Functional images sensitive to the blood oxygen level-dependent (BOLD) response were obtained by a single-shot echo-planar imaging sequence (TR 2 s, TE 48 ms, flip angle 80°,  $64 \times 64$  at 3 mm in-plane resolution, 7 mm thickness, 20 contiguous oblique axial slices parallel to the AC-PC line, 51 volumes for each run). After the experimental run, we collected anatomical images using a standard procedure (isotropic voxel size  $1 \text{ mm}^3$ ).

**Data analysis.** Images were preprocessed using SPM2 software (Wellcome Department of Imaging Neuroscience, London, UK). The six initial images of each run were discarded from the analysis to achieve steady-state equilibrium between radio-frequency pulsing and relaxation. Images were calibrated for correction of slice-acquisition timing and realigned for correction of head movement. The movement of  $x$ ,  $y$ , and  $z$  axes was  $< 1 \text{ mm}$  within each run. Following the realignment, the images were coregistered to the anatomical images, normalized to the Montreal Neurological Institute (MNI) template with affine registration, and spatially smoothed with a three-dimensional Gaussian kernel of 6 mm at full width at half maximum.

We excluded the initial response for each run corresponding to the build-up from individual analyses. Recent psychophysical studies have demonstrated that the duration of the first coherent stream is longer than that of subsequent ones (Denham and Winkler, 2006; Pressnitzer and Hupé, 2006; Kashino et al., 2007). Thus, there is the possibility that the build-up is a special state of auditory streaming. For an event-related fMRI analysis, we estimated the median reaction times (RTs) of each participant for tone detection. In all of the analyses reported in this study, the onset of all perceptual switches was temporally shifted to compensate for delays in pressing the button. We performed a group analysis with a random-effects model.

First, using a canonical hemodynamic response function (HRF), we clarified transient activations during perceptual switching and tone detection at  $\Delta f \approx 2$  and 6 semitones. We used a high-pass filter with a cutoff at 128 s to remove baseline drifts and used an autoregressive model of order one to control for temporal autocorrelation. Activations in several brain areas were significant for perceptual switching and tone detection at the threshold of FDR  $< 0.05$ . Similarly, we obtained contrasts of perceptual switching versus tone detection using one-sample  $t$  tests, but did not find differences in the extent of activations between the two tasks. To compare the intensity of activations between perceptual switching and tone detection, we extracted signal changes from local maxima of activated brain areas (see Table 1). We computed peak amplitudes of signal changes (after  $\sim 6$  s of perceptual switching and tone detection) and averaged them across all participants.

Second, using a canonical HRF and its temporal derivatives, we examined which brain areas induce perceptual switching from S1 to S2 and in the opposite direction. The group analysis results demonstrated that the posterior insular cortex (PIC), auditory cortex (AC), and medial geniculate body (MGB) in the thalamus were activated bilaterally, regardless of switching directions. We selected those areas as regions of interest (ROIs) to focus on early stages for the online analysis of acoustic features. Without anatomical normalization and spatial smoothing, we identified accurate and robust activations for each participant. The results of individual analyses showed that activations of ROIs reached a significant level in all participants ( $p < 0.001$ , uncorrected) (supplemental Fig. S1, available at [www.jneurosci.org](http://www.jneurosci.org) as supplemental material). We calculated parameter estimates of a canonical HRF and its temporal derivatives to compare temporal precedence of activations between the two directions of perceptual switching. Parameter estimates reflect the degree of the contribution of each regressor to fMRI data (Friston et al., 1998; Henson et al., 2002). For BOLD responses earlier than the canonical, parameter estimates of the derivative were positive. The temporal precedence ( $\Delta t$ ) of the response relative to the canonical HRF was computed as the ratio of derivative to canonical parameter estimates ( $\beta_2/\beta_1$ ). A positive value of the ratio indicates that the response occurs earlier than the canonical (i.e., negative value of  $\Delta t$ ). The ratio can be approximately transformed into  $\Delta t$  in seconds with the sigmoidal logistic function  $\Delta t \approx 2C/[1 + \exp(D\beta_2/\beta_1)] - C$ , where  $C = 1.78$  and  $D = 3.10$ . We analyzed our data with a different approach to validate the asymmetry of temporal precedence between

**Table 1. Activated brain areas during perceptual switching and tone detection**

Brain area	BA	Perceptual switching				Tone detection			
		x	y	z	Z score	x	y	z	Z score
$\Delta f$ ; 2 semitones ( $n = 12$ )									
Posterior insular cortex	L13	-42	12	-4	4.22	-48	12	-6	5.52
	R13	48	16	-4	4.40	46	10	-2	4.55
Auditory cortex	L41/42	-54	-24	8	4.31	-46	-16	10	5.27
	R41/42	50	-26	8	3.78	48	-14	8	5.87
Medial geniculate body	L	-16	-24	0	4.57	-16	-24	0	4.67
	R	16	-26	0	4.38	14	-24	0	5.03
Supramarginal gyrus	L40	-54	-42	32	4.31	-52	-42	32	5.50
	R40	50	-40	26	4.53	54	-44	28	5.09
Cerebellum	L	-16	-60	-24	4.48	-22	-64	-28	4.78
$\Delta f$ ; 6 semitones ( $n = 12$ )									
Posterior insular cortex	L13	-40	10	-2	3.89	-44	8	-2	4.58
	R13	-42	10	-4	4.99	46	6	0	5.42
Auditory cortex	L41/42	-52	-22	4	3.83	-50	-16	6	4.09
	R41/42	48	-18	10	4.39	48	-20	8	4.78
Medial geniculate body	L	-18	-24	0	3.56	-16	-22	0	4.33
	R	16	-22	0	3.36	20	-24	0	4.03
Cerebellum	L	-18	-64	-30	4.58	-20	-68	-28	4.98

Coordinates (x, y, z) indicate the voxel of local maxima in each area. BA, Brodmann area; L, left; R, right.

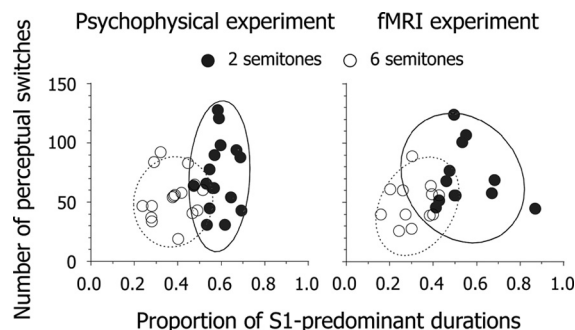
the MGB and AC activations (see supplemental data, available at [www.jneurosci.org](http://www.jneurosci.org) as supplemental material).

## Results

### Psychophysical results

We conducted a psychophysical experiment outside the scanner to select participants for the fMRI experiment. Sixteen participants were assigned the perceptual switching task at  $\Delta f \approx 2$  semitones, and the other sixteen were assigned the same task at  $\Delta f \approx 6$  semitones. In this task, the initial response for all participants was S1, which changed to S2 within several seconds. As described above, the initial switch for each run was removed from subsequent analyses. The median intervals between perceptual switching were  $5060 \pm 540$  ms (mean  $\pm$  SEM) at  $\Delta f \approx 2$  semitones and  $6630 \pm 760$  ms at  $\Delta f \approx 6$  semitones. Using logarithmic transformation, we calibrated each interval before averaging because the distribution of the intervals showed a positive skewness under the two  $\Delta f$  conditions. The transformed interval was shorter at  $\Delta f \approx 2$  semitones ( $3.66 \pm 0.05$ ) than at  $\Delta f \approx 6$  semitones ( $3.80 \pm 0.05$ ) ( $t_{(30)} = 2.38, p < 0.05$ ). We removed the duration of the first S1 from that of each run and estimated the proportion of S1-predominant durations for each participant. The results indicate that predominant percepts were S1 and S2 at  $\Delta f \approx 2$  and 6 semitones, respectively (Fig. 2).

Taking the time constants of the BOLD response into account, we recruited for the fMRI experiment 24 participants who had exhibited appropriately long intervals between perceptual switching in the psychophysical experiment. Twelve were assigned the perceptual switching and tone detection tasks at  $\Delta f \approx 2$  semitones and the other 12 were assigned the same tasks at  $\Delta f \approx 6$  semitones. The pattern of perceptual predominance in the fMRI experiment was similar to that in the psychophysical experiment (Fig. 2). The intervals between perceptual switching were  $5200 \pm 500$  ms at  $\Delta f \approx 2$  semitones and  $7210 \pm 540$  ms at  $\Delta f \approx 6$  semitones. The transformed interval was shorter at  $\Delta f \approx 2$  semitones ( $3.69 \pm 0.04$ ) than at  $\Delta f \approx 6$  semitones ( $3.84 \pm 0.04$ ) ( $t_{(22)} = 2.56, p < 0.05$ ). The accuracy of tone detection was 100% for all participants. RTs for tone detection were  $310 \pm 20$  ms and  $340 \pm 30$  ms at  $\Delta f \approx 2$  and 6 semitones, respectively. There was no difference in transformed RTs between the two  $\Delta f$  conditions:  $2.49 \pm 0.03$  and  $2.51 \pm 0.03$ .



**Figure 2.** Scatter plots of the number of perceptual switches within five 90 s runs as a function of the proportion of S1-predominant durations. Circles indicate individual data derived from psychophysical ( $N = 32$ ) and fMRI ( $N = 24$ ) experiments. Solid and dashed curved lines represent probability ellipses (90% confidence intervals) based on Mahalanobis distances at  $\Delta f \approx 2$  and 6 semitones, respectively.

Thus, audio-motor coordination for the button press did not differ between the two groups of participants.

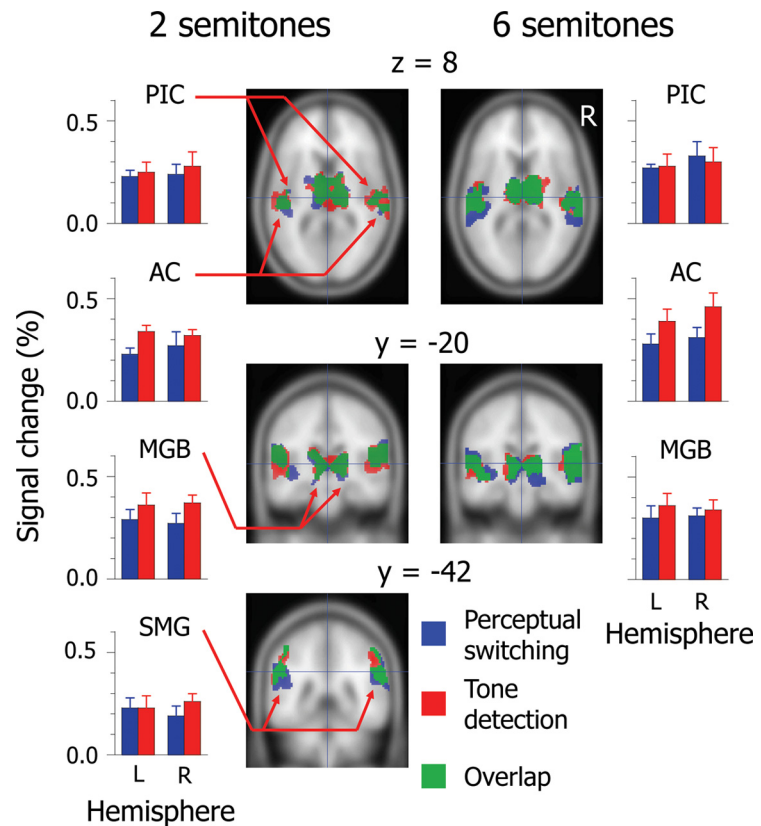
### Neuroimaging results

We identified transient activations related to perceptual switching and tone detection. The PIC, AC, and MGB were activated bilaterally at  $\Delta f \approx 6$  semitones, whereas the supramarginal gyrus (SMG), in addition to those areas, was activated at  $\Delta f \approx 2$  semitones (Table 1). We also found activation in the left cerebellum related to motor response planning. It should be noted that the results of an event-related analysis reflect activations time-locked with perceptual switching and tone detection rather than those derived from constant stimulation and background noises. Activation areas during perceptual switching overlapped those during tone detection, but amplitudes of signal changes differed between the two tasks (Fig. 3). We performed a repeated-measures ANOVA of amplitudes of signal changes. The ANOVA included the following factors: tasks (perceptual switching and tone detection); brain areas (PIC, AC, MGB, and SMG for  $\Delta f \approx 2$  semitones; PIC, AC, and MGB for  $\Delta f \approx 6$  semitones); hemispheres (left and right). At  $\Delta f \approx 2$  semitones, the amplitude of signal changes was greater for tone detection ( $0.30 \pm 0.02$ ) than for perceptual switching ( $0.24 \pm 0.03$ ) ( $F_{(1,11)} = 8.33, p < 0.05$ ).

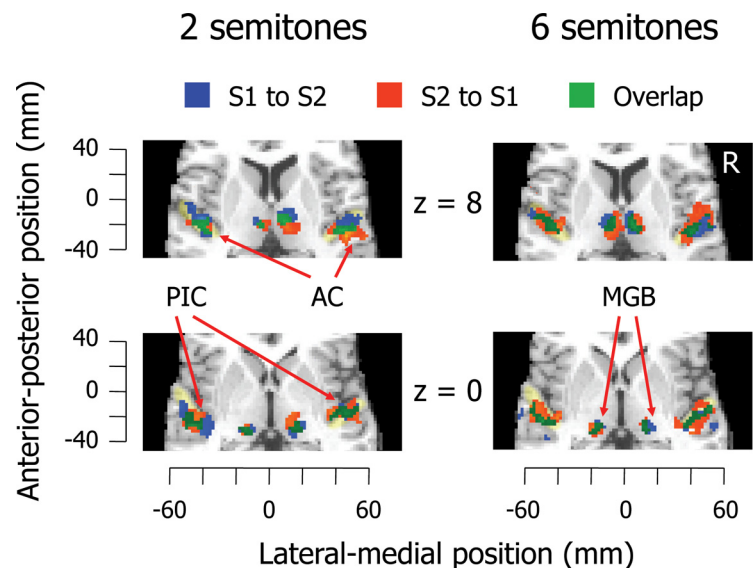
At  $\Delta f \approx 6$  semitones, there was a marginal significant difference in the amplitude of signal changes between perceptual switching ( $0.30 \pm 0.03$ ) and tone detection ( $0.35 \pm 0.04$ ) ( $F_{(1,11)} = 3.34$ ,  $p < 0.10$ ). The results indicate that the auditory system is responsible for perceptual switching as well as for tone detection, but the formation of percepts in auditory streaming is not as salient an event as the detection of deviation in constant stimulation.

Intriguingly, local maxima in the AC during perceptual switching were located at more lateral and posterior areas than those during tone detection. Local maxima observed in this study were located near dipole sources ( $x, y, z = -48, -21, 6; 52, -15, 8$ ) identified in an MEG study (Gutschalk et al., 2005), which argues that S1 and S2 percepts are maintained within nonprimary auditory areas. These findings suggest that the “formation” of percepts shares auditory functions at the level of the AC with the “maintenance” of percepts. However, in addition to AC activation, we found PIC and MGB activations. Thus, there is the possibility that higher and lower levels of the auditory system are involved in the formation of percepts in auditory streaming.

To examine which brain areas induce perceptual switching, we focused on switching directions and computed temporal precedence of activations using a canonical HRF and its temporal derivatives. We classified all events into perceptual switches from S1 to S2 and in the opposite direction. Averaged activations were significant in the PIC, AC, and MGB, but activations of other brain areas did not remain (Fig. 4). We determined the PIC, AC, and MGB as ROIs for each participant (supplemental Fig. S1, available at [www.jneurosci.org](http://www.jneurosci.org) as supplemental material) and captured timing variations in BOLD responses of ROIs. We computed correlations between the proportion of S1-predominant durations and temporal precedence of activations in ROIs. If activations during perceptual switching rely on S1 and S2 percepts, we should find consistently positive (or negative) brain–behavior correlations, regardless of switching directions. However, although correlation coefficients reached a satisfactory significant level in the AC and MGB ( $|r_{(22)}| \geq 0.28$ ,  $p < 0.20$ ), we found an interaction of correlations between the two directions of perceptual switching (Fig. 5). Specifically, the AC was activated earlier during perceptual switching from S1 to S2 as the proportion of S1-predominant durations increased. Conversely, the AC was activated earlier during perceptual switching from S2 to S1 as the proportion of S1-predominant durations decreased. However,

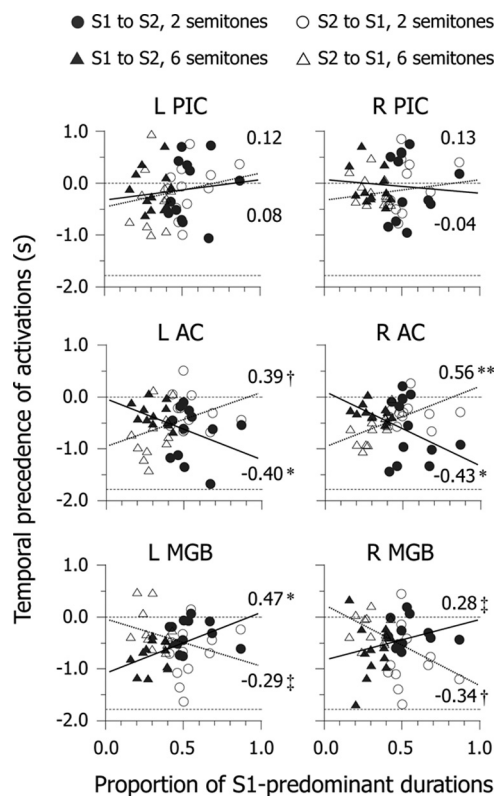


**Figure 3.** Extent and intensity of averaged activations for perceptual switching and tone detection ( $n = 12$  for each  $\Delta f$  condition). Colored areas on the coronal and horizontal planes of the standard brain indicate significant activations ( $FDR < 0.05$ ). Hairlines in top/middle and bottom panels target the coordinates  $(x, y, z) = (0, -20, 8)$  and  $(0, -42, 30)$ , respectively. The amplitudes of signal changes are derived from responses of local maxima in activated brain areas. See Table 1 for the locations of local maxima. Error bars represent the SEM. PIC, Posterior insular cortex; AC, auditory cortex; MGB, medial geniculate body; SMG, supramarginal gyrus; L, left; R, right.



**Figure 4.** Activations related to the two directions of perceptual switching. Significant activations are superimposed on the horizontal plane of the standard brain ( $p < 0.001$ , uncorrected). Areas colored in light yellow indicate Heschl gyrus. See the legends of Figures 1 and 3 for abbreviations.

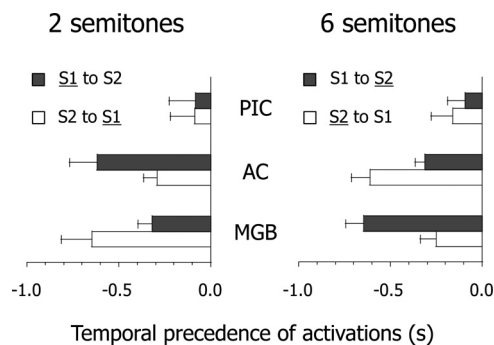
the pattern of brain–behavior correlations in the MGB was the inverse of that in the AC. These results can be summarized as follows: (1) activations in the AC and MGB, rather than in the PIC, are strongly linked with individual differences in perceptual



**Figure 5.** Scatter plots of temporal precedence of brain activity as a function of the proportion of S1-predominant durations. Circles and triangles indicate individual data ( $N = 24$  for each). Values on the ordinate axis indicate the difference in latency to the peak between an observed response and canonical hemodynamic response. Dashed lines at 0 s and  $-1.78$  s of temporal precedence represent the equality of the canonical hemodynamic response and the lower (i.e., earlier) limitation of range, respectively. Solid and dotted lines are regression lines corresponding to perceptual switching from S1 to S2 and in the opposite direction, respectively. Values in each panel represent brain–behavior correlation coefficients. See the legends of Figures 1 and 3 for abbreviations. \*\* $p < 0.01$ , \* $p < 0.05$ , † $p < 0.10$ , ‡ $p < 0.20$ .

predominance in auditory streaming; (2) the temporal precedence of AC and MGB activations does not depend on S1 and S2 percepts, but varies with switching directions; and (3) from the perspective of switching directions, the temporal precedence of AC activations is dissociated from that of MGB activation.

For ease of interpretation, we averaged the temporal precedence of activations for each  $\Delta f$  condition. Since we did not find functional lateralization of brain–behavior correlations between the two hemispheres (Fig. 5), we collapsed the data for each participant between left and right ROIs. If activations during switching are triggered by feedforward processes in a lower-level area, thalamic activation should precede cortical activation. Conversely, if activations during switching are caused by feedback processes in a higher-level area, cortical activation should precede thalamic activation. We performed a mixed-design ANOVA of the temporal precedence of activations, which included 2 ( $\Delta f$ )  $\times$  2 (switching directions)  $\times$  3 (ROIs) factors. The results showed that the three-way interaction was significant ( $F_{(2,44)} = 17.55$ ,  $p < 0.01$ ) (Fig. 6). The pattern of temporal precedence was AC (approximately  $-610$  ms)  $>$  MGB (approximately  $-290$  ms)  $>$  PIC (approximately  $-120$  ms) during switching from S1 to S2 at  $\Delta f \approx 2$  semitones and during switching from S2 to S1 at  $\Delta f \approx 6$  semitones (underline indicates the predominant percept). In contrast, it was MGB (approximately  $-640$  ms)  $>$  AC (approximately  $-300$  ms)  $>$  PIC (approximately  $-90$  ms) during



**Figure 6.** Averaged temporal precedence of brain activity. Temporal precedence is estimated by using the ratio of parameter estimates for a canonical hemodynamic response function and its temporal derivative and is collapsed between left and right brain areas for each participant. Error bars represent the SEM. Predominant percepts are underlined. See the legends of Figures 1 and 3 for abbreviations.

switching from S2 to S1 at  $\Delta f \approx 2$  semitones and during switching from S1 to S2 at  $\Delta f \approx 6$  semitones. The main point is that the response in the MGB occurs earlier during switching from non-predominant to predominant percepts, whereas that in the AC occurs earlier during switching from predominant to nonpredominant percepts. Furthermore, we confirmed the temporal dynamics of AC and MGB activations using Fourier basis functions (supplemental Fig. S2, available at [www.jneurosci.org](http://www.jneurosci.org) as supplemental material) and structural equation modeling (supplemental Fig. S3 and Table S1, available at [www.jneurosci.org](http://www.jneurosci.org) as supplemental material). Together, the results indicate that the asymmetry of temporal precedence can be explained by the assumption of an interaction between feedforward and feedback processes in the thalamocortical loop.

## Discussion

The present results demonstrate differences in brain activity during spontaneous switches in auditory streaming percepts. We focused on early stages of auditory processing to examine whether temporal dynamics of activations differs between the two directions of perceptual switching. We found that the temporal precedence of MGB and AC activations during perceptual switching is affected by perceptual predominance, regardless of switching directions. This suggests that feedforward and feedback processes in the thalamocortical loop are involved in the formation of percepts in auditory streaming.

We found that AC activations overlapped between perceptual switching and tone detection (Table 1). The pattern of the results is consistent with previous findings of fMRI, MEG, and ERP studies. First, it has been shown that perceptual representations of auditory streams are maintained in the AC of humans (Micheyl et al., 2007). As  $\Delta f$  between two pure tones (a repeating LHLH pattern) and the difference of the fundamental frequency ( $\Delta f_0$ ) between two complex tones (a repeating HLLL pattern) increases, amplitudes of AC activation increase with changes in the predominant percept from S1 to S2 (Gutschalk et al., 2007; Wilson et al., 2007). Second, the tone detection task used in the present study was postulated as an oddball one, in which mismatch negativity (MMN) (latency range of 100–250 ms) is commonly elicited and its major source is assumed to be located in the AC (Giard et al., 1990; Carlyon, 2004). Together, these findings indicate that the AC is responsible for the detection of both perceptual and physical changes in constant stimulation.

Our results indicate that, although the amplitudes of signal changes in ROIs are greater for tone detection than for perceptual switching (Fig. 3), auditory awareness emerges without changes

in physical inputs. Gutschalk et al. (2008), using informational masking, in which a stream of repeating target tones was embedded in a stochastic tone background, have revealed that an MEG response related to auditory awareness can be discriminated from one related to physical inputs. Specifically, a long-latency response (50–250 ms) was evoked in the AC when participants detected targets, but was not when they did not detect targets. In contrast, a middle-latency steady-state response was elicited when targets were detected and undetected. Thus, the neural correlates of auditory awareness might be separate from those of detection of physical changes.

More importantly, the present study showed that both MGB and AC activations are responsible for the formation of percepts in auditory streaming (Figs. 4, 5). This makes sense in terms of anatomical connections in the auditory system because the auditory system consists of ascending and descending pathways between the MGB and AC (Scott and Johnsrude, 2003; Winer et al., 2005). In monkeys, inputs from the ventral division of the MGB are projected to layers III and IV in the primary auditory area (core region), whereas those from the dorsal division of the MGB are projected to layers I, II, V, and VI in the secondary auditory area (belt and parabelt regions) (Winer, 1985; Villa et al., 1991). The MGB receives feedback from layers V and VI of the AC (Prieto and Winer, 1999; Winer and Prieto, 2001). The structure of the auditory system in humans appears to be similar to that in monkeys (Rivier and Clarke, 1997; Rademacher et al., 2001; Sweet et al., 2005). Thus, reciprocal connections between the MGB and AC appear to play an important role in perceptual switches between S1 and S2 percepts.

We found that the response in the MGB occurred earlier during switching to predominant percepts, whereas that in the AC occurred earlier during switching to nonpredominant percepts, regardless of  $\Delta f$  (Fig. 6; supplemental Fig. S2, available at [www.jneurosci.org](http://www.jneurosci.org) as supplemental material). Thus, a framework of feedforward and feedback processes between the MGB and AC is a promising way to account for the asymmetry of temporal precedence. Several researchers, particularly in visual science, have pointed out that a functional interaction between two brain areas is closely linked with alternations between bistable percepts. Tong et al. (1998) demonstrated alternations of activity between the fusiform face area and parahippocampal place area during binocular rivalry, in which monocular images of a face and house were presented to different eyes. In binocular rivalry for simple stimuli, synchronization of activations between the lateral geniculate nucleus (LGN) and primary visual area (V1) is observed during perceptual alternations. Activations in the LGN and V1 increased when participants perceived a high-contrast grating and decreased when they perceived a low-contrast grating (Wunderlich et al., 2005). Further, those activations were enhanced when their preferred-eye stimulus was perceptually dominant and were reduced when their preferred-eye stimulus was perceptually suppressed (Haynes et al., 2005). However, previous studies have mainly focused on the closer relationship between thalamocortical feedforward processes and visual awareness; they did not take corticothalamic feedback processes into consideration.

Our results showed that activations related to perceptual formation in auditory streaming depend on perceptual predominance and that nonpredominant percepts emerge from corticothalamic feedback processes (supplemental Fig. S3 and Table S1, available at [www.jneurosci.org](http://www.jneurosci.org) as supplemental material). Consider here the importance of feedback processes in the formation of percepts. Hupé et al. (1998) have found that V5 inactivation by cooling

leads to a reduction of neuronal responses in V1, V2, and V3 when anesthetized monkeys are presented with a moving bar on a stationary background. Strong effects on the response reduction in V3 were observed, particularly for low salience of the bar. Given that feedback processes in the visual cortex are recruited for low-salience stimuli, that is, perceptually “difficult” stimuli, the temporal precedence of cortical activation during switching to nonpredominant percepts identified in the present study would be consistent with the findings of Hupé et al. (1998). Furthermore, feedback connections from the visual cortex appear to optimize the contribution of thalamic functions to both local segmentation and global integration of events (Sillito and Jones, 2002). Thus, in the light of our results, we can expect that feedback processes from the AC to the MGB enhance perceptual predominance of S1 and S2 percepts.

We have demonstrated that feedforward and feedback processes between the MGB and AC are responsible for perceptual formation in auditory streaming. In addition, previous studies have revealed that auditory perceptual organization is influenced by factors of sensory adaptation (Micheyl et al., 2005; Pressnitzer et al., 2008), temporal coherence (Elhilali et al., 2009), and endogenous attention (Carlyon et al., 2001; Cusack et al., 2004). Specifically, endogenous attention appears to affect early and late stages of auditory streaming (Snyder and Alain, 2007). An fMRI study has shown that the S2 percept, rather than the S1 percept, requires more supramodal attentional demands related to activity of the intraparietal sulcus (Cusack, 2005). An ERP study has revealed that the  $N_{1c}$  response (150–250 ms) is evoked in the Heschl gyrus when participants pay attention to a sequence of repeating LHL sounds but that it is not when they ignore it (Snyder et al., 2006). For an integrative understanding of those findings, future studies should clarify brain activity at which the levels of the auditory systems are influenced by attentional manipulations. The feedforward and feedback interactions identified in the present study may be a useful concept for such investigations.

## References

- Alain C, Cortese F, Picton TW (1998) Event-related activity associated with auditory pattern processing. *Neuroreport* 9:3537–3541.
- Carlyon RP (2004) How the brain separates sounds. *Trends Cogn Sci* 8:465–471.
- Carlyon RP, Cusack R, Foxton JM, Robertson IH (2001) Effects of attention and unilateral neglect on auditory stream segregation. *J Exp Psychol Hum Percept Perform* 27:115–127.
- Cusack R (2005) The intraparietal sulcus and perceptual organization. *J Cogn Neurosci* 17:641–651.
- Cusack R, Deeks J, Aikman G, Carlyon RP (2004) Effects of location, frequency region and time course of selective attention on auditory scene analysis. *J Exp Psychol Hum Percept Perform* 30:643–656.
- Denham SL, Winkler I (2006) The role of predictive models in the formation of auditory streams. *J Physiol Paris* 100:154–170.
- Elhilali M, Ma L, Micheyl C, Oxenham AJ, Shamma SA (2009) Temporal coherence in the perceptual organization and cortical representation of auditory scenes. *Neuron* 61:317–329.
- Fishman YI, Reser DH, Arezzo JC, Steinschneider M (2001) Neural correlates of auditory stream segregation in primary auditory cortex of the awake monkey. *Hear Res* 151:167–187.
- Fishman YI, Arezzo JC, 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.
- Friston KJ, Fletcher P, Josephs O, Holmes A, Rugg MD, Turner R (1998) Event-related fMRI: characterizing differential responses. *Neuroimage* 7:30–40.
- Giard MH, Perrin F, Pernier J, Bouchet P (1990) Brain generators implicated in the processing of auditory stimulus deviance: a topographic event-related potential study. *Psychophysiology* 27:627–640.
- Gutschalk A, Micheyl C, Melcher JR, Rupp A, Scherg M, Oxenham AJ (2005)

- Neuromagnetic correlates of streaming in human auditory cortex. *J Neurosci* 25:5382–5388.
- Gutschalk A, Oxenham AJ, Micheyl C, Wilson EC, Melcher JR (2007) Human cortical activity during streaming without spectral cues suggests a general neural substrate for auditory stream segregation. *J Neurosci* 27:13074–13081.
- Gutschalk A, Micheyl C, Oxenham AJ (2008) Neural correlates of auditory perceptual awareness under informational masking. *PLoS Biol* 6:e138.
- Haynes JD, Deichmann R, Rees G (2005) Eye-specific effects of binocular rivalry in the human lateral geniculate nucleus. *Nature* 438:496–499.
- Henson RNA, Price CJ, Rugg MD, Turner R, Friston KJ (2002) Detecting latency differences in event-related BOLD responses: application to words versus nonwords and initial versus repeated face presentations. *Neuroimage* 15:83–97.
- Hupé JM, James AC, Payne BR, Lomber SG, Girard P, Bullier J (1998) Cortical feedback improves discrimination between figure and background by V1, V2 and V3 neurons. *Nature* 394:784–787.
- Kashino M, Okada M, Mizutani S, Davis P, Kondo HM (2007) The dynamics of auditory streaming: psychophysics, neuroimaging, and modeling. In: *Hearing—from sensory processing to perception* (Kollmeier B, Klump G, Hohmann V, Langemann U, Mauermann M, Uppenkamp S, Verhey J, eds), pp 275–283. Berlin: Springer.
- Micheyl C, Tian B, Carlyon RP, Rauschecker JP (2005) Perceptual organization of tone sequences in the auditory cortex of awake macaques. *Neuron* 48:139–148.
- Micheyl C, Carlyon RP, Gutschalk A, Melcher JR, Oxenham AJ, Rauschecker JP, Tian B, Wilson EC (2007) The role of auditory cortex in the formation of auditory streams. *Hear Res* 229:116–131.
- Moore BCJ, Gockel H (2002) Factors influencing sequential stream segregation. *Acta Acust United Acust* 88:320–332.
- Näätänen R, Tervaniemi M, Sussman E, Paavilainen P, Winkler I (2001) 'Primitive intelligence' in the auditory cortex. *Trends Neurosci* 24: 283–288.
- Pressnitzer D, Hupé JM (2006) Temporal dynamics of auditory and visual bistability reveal common principles of perceptual organization. *Curr Biol* 16:1351–1357.
- Pressnitzer D, Sayles M, Micheyl C, Winter IM (2008) Perceptual organization of sound begins in the auditory periphery. *Curr Biol* 18:1124–1128.
- Prieto JJ, Winer JA (1999) Layer VI in cat primary auditory cortex: Golgi study and sublaminal origins of projection neurons. *J Comp Neurol* 404:332–358.
- Rademacher J, Morosan P, Schormann T, Schleicher A, Werner C, Freund HJ, Zilles K (2001) Probabilistic mapping and volume measurement of human primary auditory cortex. *Neuroimage* 13:669–683.
- 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.
- Scott SK, Johnsrude IS (2003) The neuroanatomical and functional organization of speech perception. *Trends Neurosci* 26:100–107.
- Sillito AM, Jones HE (2002) Corticothalamic interactions in the transfer of visual information. *Philos Trans R Soc Lond B Biol Sci* 357:1739–1752.
- Snyder JS, Alain C (2007) Toward a neurophysiological theory of auditory stream segregation. *Psychol Bull* 133:780–799.
- Snyder JS, Alain C, Picton TW (2006) Effects of attention on neuroelectric correlates of auditory stream segregation. *J Cogn Neurosci* 18:1–13.
- Sussman E, Ritter W, Vaughan HG Jr (1999) An investigation of the auditory streaming effect using event-related brain potentials. *Psychophysiology* 36:22–34.
- Sweet RA, Dorph-Petersen KA, Lewis DA (2005) Mapping auditory core, lateral belt, and parabelt cortices in the human superior temporal gyrus. *J Comp Neurol* 491:270–289.
- Tong F, Nakayama K, Vaughan JT, Kanwisher N (1998) Binocular rivalry and visual awareness in human extrastriate cortex. *Neuron* 21:753–759.
- van Noorden LPAS (1975) Temporal coherence in the perception of tone sequences. PhD thesis, Eindhoven University of Technology.
- Villa AE, Rouiller EM, Simm GM, Zurita P, de Ribaupierre Y, de Ribaupierre F (1991) Corticofugal modulation of the information processing in the auditory thalamus of the cat. *Exp Brain Res* 86:506–517.
- Wilson EC, Melcher JR, Micheyl C, Gutschalk A, Oxenham AJ (2007) Cortical fMRI activation to sequences of tones alternating in frequency: relationship to perceived rate and streaming. *J Neurophysiol* 97:2230–2238.
- Winer JA (1985) Structure of layer II in cat primary auditory cortex (AI). *J Comp Neurol* 238:10–37.
- Winer JA, Prieto JJ (2001) Layer V in cat primary auditory cortex (AI): cellular architecture and identification of projection neurons. *J Comp Neurol* 434:379–412.
- Winer JA, Miller LM, Lee CC, Schreiner CE (2005) Auditory thalamocortical transformation: structure and function. *Trends Neurosci* 28:255–263.
- Wunderlich K, Schneider KA, Kastner S (2005) Neural correlates of binocular rivalry in the human lateral geniculate nucleus. *Nat Neurosci* 8:1595–1602.