



**HAL**  
open science

## Brain-Speech Alignment Enhances Auditory Cortical Responses and Speech Perception

Houda Saoud, Goulven Josse, Eric Bertasi, Eric Truy, Maria Chait, Anne-Lise Giraud

► **To cite this version:**

Houda Saoud, Goulven Josse, Eric Bertasi, Eric Truy, Maria Chait, et al.. Brain-Speech Alignment Enhances Auditory Cortical Responses and Speech Perception. *Journal of Neuroscience*, 2012, 32 (1), pp.275-281. 10.1523/JNEUROSCI.3970-11.2012 . hal-03994757

**HAL Id: hal-03994757**

**<https://hal.science/hal-03994757>**

Submitted on 17 Feb 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Brain–Speech Alignment Enhances Auditory Cortical Responses and Speech Perception

Houda Saoud,<sup>1,2,3</sup> Goulven Josse,<sup>1</sup> Eric Bertasi,<sup>4</sup> Eric Truy,<sup>3,5</sup> Maria Chait,<sup>6</sup> and Anne-Lise Giraud<sup>1</sup>

<sup>1</sup>Institut National de la Santé et de la Recherche Médicale (INSERM) U960, Department of Cognitive Studies, Ecole Normale Supérieure, 75005 Paris, France, <sup>2</sup>Advanced Bionics, Clinical Research Department, Great Shelford, Cambridge CB2 5LD, United Kingdom, <sup>3</sup>INSERM U1028 and UMR 5292, Lyon Neuroscience Research Center, Lyon F-69000, France, <sup>4</sup>Centre de NeuroImagerie de Recherche, Centre de Recherche de l'Institut Cerveau-Moelle (Université Pierre et Marie Curie UMR S975, CNRS UMR 7225, INSERM UMR 975), Institut du Cerveau et de la Moelle épinière, 75651 Paris Cedex 13, France, <sup>5</sup>Département d'ORL, Hôpital Edouard Herriot, 69003 Lyon, France, and <sup>6</sup>University College of London Ear Institute, London WC1X, United Kingdom

Asymmetry in auditory cortical oscillations could play a role in speech perception by fostering hemispheric triage of information across the two hemispheres. Due to this asymmetry, fast speech temporal modulations relevant for phonemic analysis could be best perceived by the left auditory cortex, while slower modulations conveying vocal and paralinguistic information would be better captured by the right one. It is unclear, however, whether and how early oscillation-based selection influences speech perception. Using a dichotic listening paradigm in human participants, where we provided different parts of the speech envelope to each ear, we show that word recognition is facilitated when the temporal properties of speech match the rhythmic properties of auditory cortices. We further show that the interaction between speech envelope and auditory cortices rhythms translates in their level of neural activity (as measured with fMRI). In the left auditory cortex, the neural activity level related to stimulus–brain rhythm interaction predicts speech perception facilitation. These data demonstrate that speech interacts with auditory cortical rhythms differently in right and left auditory cortex, and that in the latter, the interaction directly impacts speech perception performance.

## Introduction

Asymmetric sampling in time (AST) (Poehpel, 2003) could account for a selective triage of information across auditory cortices (Zatorre and Belin, 2001; Zatorre and Gandour, 2008). While dominant sampling at low gamma rate (~25–40 Hz) in left auditory cortex could facilitate the encoding of fast amplitude modulations in speech, e.g., consonant bursts and fast formant transitions, dominant sampling at theta rate (~4 Hz) in right auditory cortex would foster the encoding of slower acoustic modulations or stationary cues, e.g., vocal and prosodic signals. There has been abundant physiological evidence supporting this theory (Zaehle et al., 2004; Boemio et al., 2005; Giraud et al., 2007; Luo and Poeppel, 2007; Abrams et al., 2008; Obleser et al., 2008; Telkemeyer et al., 2009; Morillon et al., 2010), yet whether asymmetric temporal integration is directly relevant to speech perception remains unclear.

An important, and thus far lacking, piece of evidence is a psychophysical demonstration of a perceptual advantage when

each auditory cortex receives information at the temporal scale it is best tuned to. In other words, if AST theory is both correct and behaviorally relevant, we should observe a gain in perception when each auditory cortex receives sound modulations on the time scale it prefers.

We used dichotic listening in 100 participants to probe whether speech perception is facilitated when speech high- and low-rate envelopes are presented to the right and left ears, respectively, relative to the opposite dichotic presentation mode. We produced degraded speech sounds using an algorithm (Fig. 1) (Drullman et al., 1994; Chait et al., 2005) that decomposes the speech envelope into its slow (0–4 Hz) and fast (22–40 Hz) components. These specific values correspond to cortical oscillations ranges that are spontaneously predominant in right and left auditory cortices, respectively (Zatorre and Belin, 2001; Zaehle et al., 2004; Boemio et al., 2005). Given that crossed connectivity from the cochlea to the cortex is faster than ipsilateral connectivity (4 vs 5 relays due to either an additional relay in the medial superior olive or an extra transcallosal step), and that dichotic listening suppresses ipsilateral routes (Westerhausen et al., 2008), we reasoned that providing the fast temporal modulations to the right ear should induce a more direct interaction with the gamma oscillatory activity that is present at rest in the left auditory cortex (Giraud et al., 2007) than when the same acoustic information is provided to the opposite ear. Direct stimulus–brain interaction could result in a stronger reset or entrainment of intrinsic cortical oscillations (theta and low gamma), which may facilitate speech processing and subsequent detection (Schroeder and Lakatos,

Received Aug. 3, 2011; revised Oct. 21, 2011; accepted Oct. 31, 2011.

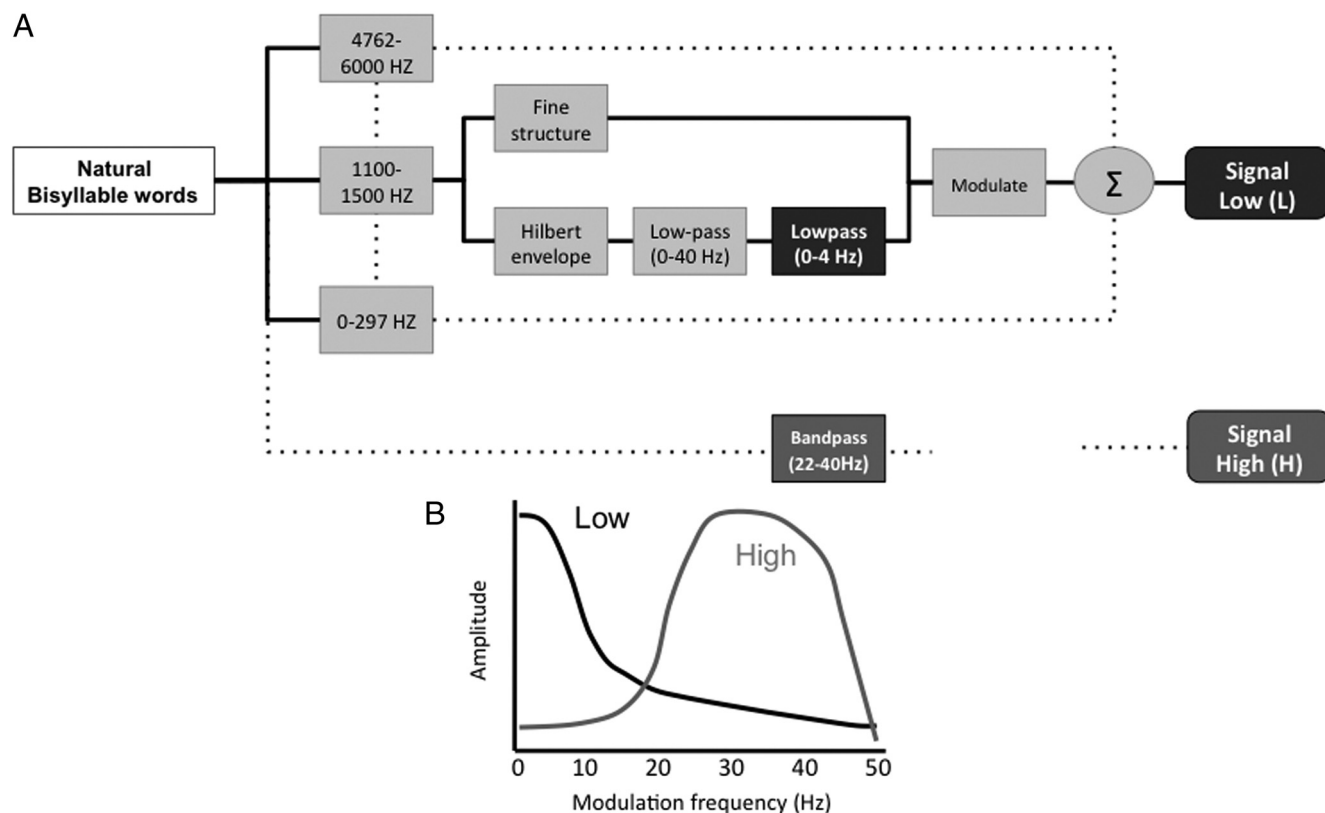
Author contributions: E.T., M.C., and A.-L.G. designed research; H.S., G.J., and E.B. performed research; H.S. and M.C. contributed unpublished reagents/analytic tools; H.S., G.J., E.B., and A.-L.G. analyzed data; H.S. and A.-L.G. wrote the paper.

This work was supported by Advanced Bionics (to H.S.) and by CNRS, INSERM, and the European Research Council (A.-L.G.). The code used to generate the stimuli was based on code developed by Takayuki Arai and Steven Greenberg.

Correspondence should be addressed to Anne-Lise Giraud, INSERM U960, Department of Cognitive Studies, Ecole Normale Supérieure, 29 rue d'Ulm, 75005 Paris, France. E-mail: anne-lise.giraud@ens.fr.

DOI:10.1523/JNEUROSCI.3970-11.2012

Copyright © 2012 the authors 0270-6474/12/320275-07\$15.00/0



**Figure 1.** Stimuli. **A**, Signal processing block diagram. Signals were low-pass filtered at 6 kHz, sampled at 16 kHz, and quantized with 16-bit resolution. The frequency spectrum of the speech signal was partitioned into 14 frequency bands with a linear-phase finite impulse response filter bank (slopes 60 dB/100 Hz or greater), spanning the range 0.1–6 kHz, spaced in 1/3 octave steps (approximately critical band-wide) across the acoustic spectrum. The Hilbert transform was used to decompose the signal in each band into a slowly varying temporal envelope and a rapidly varying fine structure. The temporal envelope was subsequently low-pass filtered with a cutoff frequency of 40 Hz and then either low- (0–4 Hz) or band- (22–40 Hz) pass filtered. The time delays, relative to the original signal, introduced by the filtering, were compensated by shifting the filter outputs. After the filtering, the envelope was combined with the carrier signal (fine structure) by multiplying the original band by the ratio between the filtered and original envelopes. The result for each original signal ( $S$ ) is  $S_{\text{low}}$  and  $S_{\text{high}}$ , containing only low- or high-modulation frequencies. **B**, Schematic representation of modulation spectrum for each stimulus category.

2009; Besle et al., 2011). We therefore hypothesized that stronger auditory cortical interaction in the preferred dichotic condition should translate into more accurate speech detection relative to the inverted dichotic presentation mode.

In addition, if dichotic stimuli match the rhythmic properties of each auditory cortex, we should observe stronger neural activity in both auditory cortices in the preferred dichotic condition than in the inverted one. We therefore further explored whether enhanced auditory activity scales with speech perception performance using fMRI.

## Materials and Methods

### Subjects

One hundred French native speakers, with no history of neurological, psychiatric, or auditory symptoms, and no specific auditory training took part in four experiments. A first psychophysics experiment (Experiment 1) included 41 subjects (mean age, 25.6 years; 20 males). Thirty-eight of them were right-handed, as measured with the Edinburgh Handedness Inventory (Oldfield, 1971). Sixteen subjects (mean age, 32.45 years; 15 males) took part in the first additional psychophysics experiment (Experiment 2), and 21 other subjects (mean age, 24.6 years; 9 males) took part in the second additional psychophysics experiment (Experiment 3). All subjects in Experiments 2 and 3 were right-handed. Twenty-two other subjects (mean age, 24.8 years; 14 right-handed, 12 males) participated in an fMRI study (Experiment 4). All experimental procedures were approved of by the local ethics committee and written informed consent was obtained from each participant.

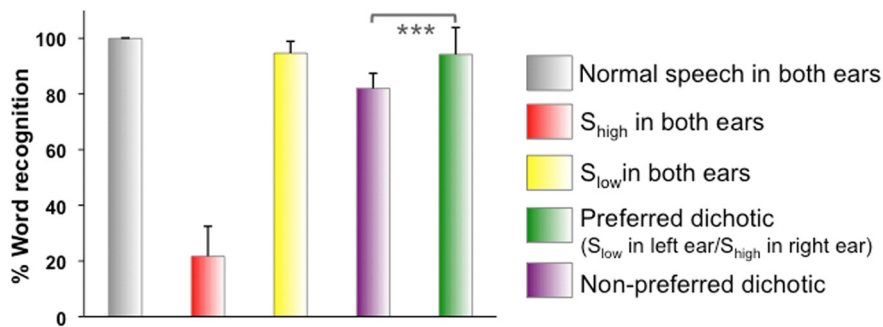
### Stimuli

The test material consisted of disyllabic French words (mean length, ~500 ms), from which we extracted and split the temporal envelope into two frequency ranges. Figure 1 describes our signal processing technique (Silipo et al., 1999), which is an extension of the method used by Drulman et al. (1994). The original wide-band speech signal was divided into 14 frequency bands with a finite impulse response filter bank, spanning the range from 0 to 6 kHz spaced in 1/3 octave steps across the acoustic spectrum. The amplitude envelope of each band was computed by means of a Hilbert transform and then either low- (0–4 Hz) or high- (22–40 Hz) bandpass filtered before being combined again with the original carrier signal. Therefore, the modified signal contains only low or fast modulation frequencies, yielding two signals called  $S_{\text{low}}$  and  $S_{\text{high}}$ , respectively.

### Experimental protocol

Before starting the tests, subjects completed the handedness questionnaire. Subjects were naive with regards to the objective of the experiment and were instructed to pay attention to what they heard in both ears. The response they gave varied depending on the experiment. The digitally recorded stimuli were presented simultaneously and binaurally at clearly audible levels (70 dB SL) through an MRConfon device (anti-noise system developed for MRI). Experiments 1, 2, and 3 were performed in a soundproof room. Experiment 4 was performed in the fMRI scanner with the scanner running continuously.

**Experiment 1.** Subjects were exposed to the processed speech stimuli. They were instructed to press the space key of a keyboard after each stimulus and to immediately repeat what they heard, even fragments of



**Figure 2.** Behavioral results of Experiment 1. Percentage of speech recognition (whole-word scoring). \*\*\* indicate *t*-test with  $p < 0.001$ .

the stimulus. We calculated the number of correctly perceived words. Importantly, such a repetition task is optimized for rating speech recognition rather than reaction times. In this first experiment, we used a total of 600 everyday phonetically balanced French words (mean length, ~500 ms), equilibrated for usage frequency. Among the 120 words of each condition, 60 were pronounced by a male speaker, and the other 60 by a female speaker. Both the original and the filtered stimuli were presented in five listening conditions: three diotic (the same signal presented in both ears) and two dichotic (a different signal presented in each ear). The five conditions were as follows: diotic: (1) Ct (control), unaltered bisyllabic words; (2)  $S_{high}L-S_{high}R$ : high-frequency envelope in both [left (L) and right (R)] ears (22–40 Hz); and (3)  $S_{low}L-S_{low}R$ : low-frequency envelope in both ears (0–4 Hz); dichotic: (4)  $S_{low}L-S_{high}R$ :  $S_{low}$  in the left ear and  $S_{high}$  frequency envelope to the right ear (this condition is the one for which we hypothesized a bilateral cortical preference); and (5)  $S_{high}L-S_{low}R$ : the inverted condition with  $S_{low}$  in the right ear and  $S_{high}$  in the left ear, referred to as nonpreferred dichotic stimulation. The order of presentation within and across conditions and subjects was randomized using MATLAB.

**Experiment 2.** The second experiment tested a new group of 16 subjects and used 400 words from Experiment 1. Experiment 2 also included new conditions as well as some changes in specific instrumentation and certain aspects of procedures.

We did not use natural stimuli in this experiment or the following ones. Stimuli were presented in eight listening conditions. Two diotic and two dichotic conditions were the same as in Experiment 1: (1)  $S_{high}$  on both ears (22–40 Hz); (2)  $S_{low}$  in both ears (0–4 Hz); (3)  $S_{low}L-S_{high}R$ : the hypothesized hemispheric preference; (4)  $S_{high}L-S_{low}R$ : non-preferred condition. Four additional dichotic conditions were included, consisting of a target stimulus in one ear and a noise in the opposite ear: (5)  $S_{low}L-NoiseR$ : envelope presented to the left ear and noise simultaneously presented to the right ear; (6)  $NoiseL-S_{low}R$ : noise to left ear and  $S_{low}$  to right ear; (7)  $S_{high}L-NoiseR$ :  $S_{high}$  presented to the left ear and noise simultaneously presented to the right ear; and (8)  $NoiseL-S_{high}R$ : noise to left ear and high envelope to right ear.

All stimuli in Experiment 2 were mixed with white noise at a signal-to-noise ratio (SNR) of 12 dB, filtered with Butterworth filter at cutoff frequency of 500 Hz. The task was similar to the task in Experiment 1, but responses were recorded using a voice-key developed under MATLAB that measured the onset of vocal response. With this change, the subjects no longer needed to press the keyboard to repeat what they heard, but were invited to focus on each stimulus and immediately repeat it aloud. As in Experiment 1, the stimuli were presented in random order. In this experiment, we rated the number of correctly repeated syllables.

**Experiment 3.** Experiment 3 used the same stimuli, conditions, and testing procedures as in Experiment 2, except that no background noise was included, and a new group of 21 subjects was tested.

**Experiment 4.** We tested 21 new subjects in Experiment 4 and used 250 stimuli mixed with white noise (SNR equal to 0 dB). Experimental conditions were as in Experiment 1. Among the 50 stimuli of each condition, 25 were pronounced by a male speaker and the other half by a female speaker. The stimuli were presented under five listening conditions, three

diotic and two dichotic: (1) Ct: unaltered bisyllabic words; (2)  $S_{high}$  in both ears (22–40 Hz); (3)  $S_{low}$  in both ears (0–6 Hz); (4)  $S_{low}L-S_{high}R$ : preferred dichotic condition; and (5)  $S_{high}L-S_{low}R$ : opposing the hemispheric preference.

The fMRI experiment consisted of two sessions: a familiarization session and a test session during which fMRI data were acquired. The familiarization procedure was similar to the one described above for Experiment 1 and comprised a set of 40 stimuli (mixed with white noise at 0 dB SNR). During the fMRI acquisition, subjects wore headphones for noise protection and delivery of acoustic stimuli. They were instructed to stay awake and as still as possible, and to press a response box as quickly as possible after hearing the stimuli. The key

controlled the display of written items on a screen viewed through a mirror. Subjects reported the word they heard by choosing between three items appearing simultaneously: the heard word, a phonological neighbor, and a question mark placed between the two words, to be selected if the subject did not understand the stimulus. This task was preferred to word repetition to minimize both motion artifacts and the effect of preparing for speaking on auditory cortex (suppression in anticipation of feedback) (Kell et al., 2011). The stimuli were divided into five sessions in which the conditions were equally represented. In each session, the order of presentation of the stimuli was randomized, as well as the interstimulus interval (between 2 and 5 s). The test session lasted ~1 h. Stimuli were presented and projected onto the screen using MATLAB and Cogent (Psychophysics Toolbox).

Imaging was performed on a 3T Siemens Trio TIM using the standard 12-channel head coil. The subjects' head movements were restrained by additional padding inside of the head coil. A high-resolution T1-weighted image was acquired at the end of the scanning (slices, 176; echo time, 4.18 ms; repetition time, 2300 ms; flip angle, 9°; pixel size, 1 × 1 × 1 mm; matrix size, 256 × 256) to exclude subjects with possible structural brain anomalies. Language brain regions were localized individually for each subject using passive listening of a short story. Functional images were acquired using a BOLD-sensitive gradient EPI using the following acquisition parameters: slices, 37; matrix size, 64 × 64; pixel size, 3 × 3 × 3 mm; echo time, 30 ms; repetition time, 2000 ms.

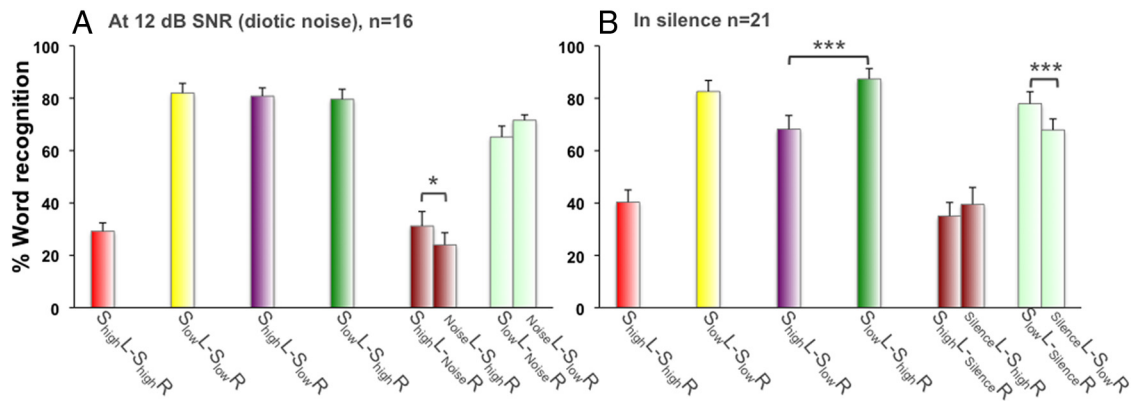
The statistical analysis of the fMRI data was performed using SPM8 and was preceded by three preprocessing steps: (1) realignment, (2) coregistration, and (3) normalization (transformation of the data into a standard atlas space). The atlas representative template conformed to the Montreal Neurological Institute atlas. Data were analyzed using the general linear model as implemented in SPM8.

## Results

### Brain–speech alignment boosts speech perception

In the first psychophysics experiment, subjects listened to processed disyllabic words in which the slow (0–4 Hz;  $S_{low}$ ) and rapid (22–40 Hz;  $S_{high}$ ) modulations were selectively extracted. In the preferred condition,  $S_{low}$  was presented to the left ear and  $S_{high}$  to the right ear, following the hypothesis that such a combination would more strongly interact with both auditory cortices than in the inverse dichotic condition (nonpreferred). Accordingly, we observed a statistically better global word recognition [94% vs 81%,  $p < 0.001$ , partial  $\eta^2 = 0.477$  (large effect  $\eta^2 > 0.1379$ ) (Cohen, 1988)] in the preferred dichotic presentation than in the nonpreferred (Fig. 2). Our results demonstrate a gain in word recognition of ~13% in the condition corresponding to the hypothesized hemispheric preference. Consistent with previous observations that speech envelopes containing only low-rate modulation frequencies can be fairly well recognized (Van Tasell et al., 1987; Drullman et al., 1994; Shannon et al., 1995; Elliott and Theunissen, 2009), speech recognition was nearly equivalent in





**Figure 3.** Behavioral results of Experiments 2 and 3. **A**, Experiment 2 in noise (12 dB SNR): percentage of word recognition (syllable scoring). The expected dichotic effect is absent (no difference between purple and dark). **B**, Experiment 3 in silence: percentage of word recognition (syllable scoring). The expected effect is present. \* and \*\*\* indicate *t* tests with  $p < 0.05$  and  $p < 0.001$ , respectively.

the  $S_{low}$  diotic condition and in the control condition where unprocessed words were presented. In the diotic  $S_{high}$  envelope condition, recognition was very low (20%). Note that in the best dichotic condition, recognition was lower than in the  $S_{low}$  diotic condition, yet above the mean of the two filtered diotic conditions. Understanding the mechanisms by which the two types of temporal information are combined in the brain is, however, outside the scope of this study. The envelope was only split as a paradigm to probe auditory cortex preference to specific modulation rates. It is important to keep in mind that even if, as hypothesized, each auditory cortex is best tuned to one scale of temporal information, it is sensitive to both and there is always a benefit from getting the two types of information from both ears.

#### Noise degrades speech–brain interaction

Although this first experiment validated our hypothesis, there were several limitations, which we sought to address in the second experiment. There was a ceiling effect in the  $S_{low}$  diotic condition and speech recognition was suboptimally measured because subjects gave their answer after pressing a key with the right hand. We therefore performed a second experiment in which we tested 16 right-handed participants using a voice-key to directly record repeated words. In this experiment, we measured syllabic accuracy rather than whole-word recognition. In addition, to eliminate the ceiling effects in the diotic  $S_{low}$  condition, we added noise to all conditions. We used a 12 dB signal-to-noise ratio after a pilot experiment aiming at obtaining the right level of perception attenuation in the diotic  $S_{low}$  condition. We added new conditions where filtered speech stimuli were only presented to one ear with noise (or silence) in the opposite ear. In the conditions with noise in the opposite ear, noise was also mixed with filtered speech signals. To limit the experiment duration (to  $\sim 1$  h), we removed the condition with unprocessed speech. To control for the change in experimental parameters, we also repeated this new experiment in yet another group of 21 subjects in silence.

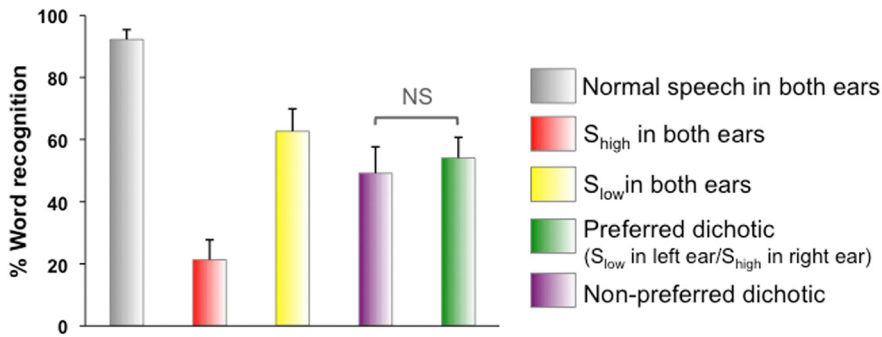
Speech recognition facilitation in the preferred relative to nonpreferred dichotic condition was reproduced in the new experiment in which no noise was added (Fig. 3A). Yet, the addition of noise abolished the hypothesized effect of presentation side. In the silent condition (Fig. 3B), the new monaural conditions confirmed the auditory cortices' preference observed in the dichotic conditions. Speech recognition was systematically better in the condition where the filtered stimulus matched the hypothesized

auditory cortex preference, i.e., when the  $S_{low}$  envelope was delivered in the left ear and the  $S_{high}$  in the right ear (*t* test,  $p < 0.001$ ,  $\eta^2 = 0.832$ , large effect). The effect was significant between the two  $S_{low}$  conditions [*t* test,  $p \leq 0.001$ ,  $\eta^2 = 0.449$  (large effect)] and close to significance for the  $S_{high}$  ones. In the experiment with noise (Fig. 3A), the effects were inverted with enhanced recognition when  $S_{low}$  stimuli were presented to the right ear and  $S_{high}$  to the right ear [*t* test,  $p = 0.032$ ,  $\eta^2 = 0.552$  (large effect) for the  $S_{high}$  comparison].

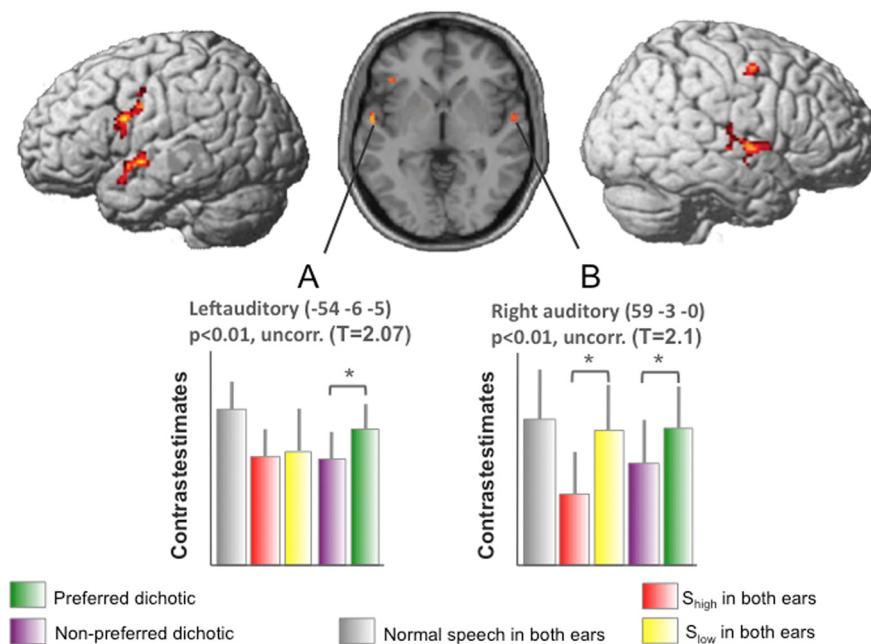
#### Brain–speech alignment enhances neural responses in auditory cortices

To assess to what extent the observed speech perception facilitation in the preferred dichotic condition could be accounted for by enhanced neural activity in auditory cortices, e.g., enhanced reset and/or entrainment, we used the same stimuli in an fMRI experiment in which we modified the task while keeping the syllabic scoring used in the second series of psychophysics experiments, hoping that the MRI scanner's structured noise would not compromise the hypothesized dichotic effects. Twenty-two new subjects (mean age, 24.8 years; 12 males; 14 right-handed) were presented with diotic, unprocessed speech, dichotic  $S_{low}$  and  $S_{high}$  speech signals in the preferred and nonpreferred lateralization settings, and diotic  $S_{low}$  and  $S_{high}$  signals. In each of those five conditions, participants had to press a key as soon as they heard the word. The key controlled the display of a screen showing two words (phonological neighbors, e.g., poison, poisson) that appeared after a variable (randomized) delay to ensure statistical separability of the auditory and visual events. Subjects had to choose which of the two words corresponded to the one they heard using another keypress. We analyzed the fMRI data using a GLM in which we modeled the sound onset, the keypress, the screen presentation, and the final response. We focused our analysis on the contrasts between the two preferred and nonpreferred dichotic sound conditions. As our sample also included left-handers, we analyzed the effect of handedness behaviorally and further factored it out in the GLM, even though it did not yield significant effect in auditory cortices.

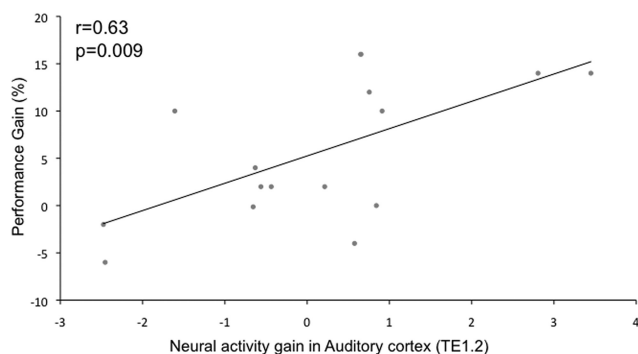
The behavioral data obtained while scanning were qualitatively consistent with our previous psychophysical observations (Fig. 4). In the preferred dichotic condition, where each auditory cortex received the part of the envelope they were supposedly most sensitive to, we observed a gain in speech recognition, as measured by choosing between phonological neighbors. Yet,



**Figure 4.** Behavioral results for the fMRI in Experiment 4. Percentage of speech recognition (syllable scoring). The expected dichotic effect is qualitatively present (nonsignificant).



**Figure 5.** Whole-brain fMRI results. The preferred dichotic condition yielded more neural activity in bilateral auditory and premotor cortices. The effects remained significant after correction for multiple comparisons within the cytoarchitectonic map of auditory cortices (our hypothesis; \* indicate  $p < 0.01$  after correction), but no longer across the whole brain. Bottom, Relative level of activity across conditions in left auditory cortex (**A**) and right auditory cortex (**B**) in 1 cm spheres centered around the coordinates in parentheses.



**Figure 6.** Positive correlation between neural activity gain measured from cytoarchitectonic region TE1.2 (fMRI experiment,  $S_{high}$  in left ear/ $S_{low}$  in right ear minus  $S_{low}$  in left ear/ $S_{high}$  in right ear) and performance gain (difference in speech recognition score between the two dichotic conditions).

these effects were far from significant ( $t$  tests,  $p = 0.292$ ). Presumably, the scanner noise interfered with the expected effect in a similar way as white noise did in the second psychophysics experiment. Interestingly however, despite the lack of behavioral effects, we observed significantly enhanced neural responses in bilateral auditory cortices in the preferred dichotic conditions relative to the other one. These effects remained below significance when corrected for multiple comparisons on a whole-brain basis ( $p < 0.005$ , uncorrected; Fig. 5), but were statistically significant when considering a region of interest centered on auditory cortex (probabilistic map of auditory cortex,  $p < 0.01$ , corrected). The effect localized to the TE1.2 part of human auditory cortex (Morosan et al., 2001; Eickhoff et al., 2005, 2006), which corresponds to its anterior-most part. The left premotor articulatory region (BA44/6) was the only additional region that also showed an effect at equivalent statistical threshold.

**Increase in auditory cortex neural activity predicts speech recognition gain**

To test for a more direct prediction of behavior by neural activity, we computed correlations between auditory cortical activity and individual speech recognition scores. We found that the gain in neural activity observed in the preferred relative to nonpreferred dichotic condition positively predicted the gain in speech recognition scores ( $r = 0.63$ ,  $p < 0.01$ ; Fig. 6). There was no such correlation in the right auditory cortex (even a negative trend). Additionally, we observed a significant difference between diotic conditions ( $S_{high}$  both ears vs  $S_{low}$  both ears,  $p < 0.01$ , uncorrected;  $T = 2.8$ ) in right but not in left auditory cortex, yet no significant condition-by-side interaction (Fig. 5).

Finally, the nonpreferred condition enhanced neural activation in the right STS only. Activity in this region was selectively enhanced in the condition where the stimuli did not match the auditory cortex preference, while all other conditions elicited weak responses. In the preferred condition, we observed a very weak trend for a negative correlation between neural activity in this region and word recognition ( $r = -0.35$ ,  $p = 0.1$ ).

In sum, while behavioral gain was weak in this fMRI experiment, we found word comprehension to be correlated with neural activity of left auditory cortex.

**Discussion**

**Weak but consistent dichotic effects**

Our results across this series of psychophysics and functional neuroimaging experiments consistently indicate a preference of the left auditory cortex for stimuli that are modulated in time on a gamma frequency range, and a preference of the right auditory

cortex for those modulated on a theta frequency range. This preference manifests at the neural level by increased responses to these specific modulation rates in auditory cortices, which paralleled a performance gain only in the left auditory cortex. This observation agrees with the notion that the left auditory cortex, in particular its activity in the low gamma band, contributes more strongly to speech perception than the right auditory cortex. Presumably, if our task had emphasized paralinguistic aspects of the stimuli, e.g., gender identification, positive correlations would have been found with neural gain in right auditory cortex.

Consistent with our previous observations that premotor and auditory cortex rhythms align during linguistic stimulation (Morillon et al., 2010), enhanced neural response to the preferred dichotic stimulus was also present though unrelated to performance gain in the left inferior prefrontal cortex. Dichotic listening usually taps into prefrontal resources (Jäncke and Shah, 2002; Jäncke et al., 2003), yet the fact that prefrontal activation was unrelated to speech performance gain makes it unlikely to reflect dichotic integration processes. The present results therefore do not support a key role of temporal–prefrontal alignment in speech perception, which does not exclude that it might be essential for other tasks, i.e., speech production.

Behavioral and neural effects in this series of experiments were consistently weak and susceptible to noise. When designing these experiments, we were aware that we might not be able to observe any effect at all, despite auditory cortices sensitivity to different temporal scales, because the auditory pathways between the cochlea and the cortex are both crossed and uncrossed. Our experimental design was therefore strongly dependent on a faster contralateral cochlea–cortex interaction, which could only be experimentally useful if we used brief stimuli that could be recognized before interhemispheric cross talk and top-effects diluted behavioral effects arising from early bottom-up stimulus–brain interaction. To our advantage however, contralateral effects become predominant under dichotic listening conditions (Westerhausen et al., 2008). While the remaining ipsilateral effects might account for the weakness of our observations, the reproducibility of the latter across several experiments involving 100 participants speaks to the validity of our hypothesis. Testing whether these results would hold for connected speech remains unclear and would be an experimental challenge. It is also unclear whether we would also observe a behavioral gain with longer stimuli or connected speech due to the addition of stimulus–brain resonance at slow rate.

#### Possible influence of cochlear efferents in dichotic effects in noise

The use of noise to correct for a ceiling effect in the  $S_{low}$  condition in Experiment 1 proved to have a complex effect on speech perception. Any stimulus provided to one ear produces a suppression of outer hair cell responses in the opposite ear through a fast crossed efferent pathway involving in the superior medial olive (Guinan and Gifford, 1988; Cooper and Guinan, 2003). This suppressive effect on outer hair cells improves speech-in-noise intelligibility (Giraud et al., 1997) because it enhances the dynamic range of primary auditory neurons (Micheyl and Collet, 1996; Kumar and Vanaja, 2004). The efferent effect is known to

Condition	Cochlear efferent effect on filtered speech signal	Auditory cortex resonance with filtered speech signal	Predicted speech perception
$S_{high}L\text{-}NoiseR$	weak	weak	+
$NoiseL\text{-}S_{high}R$	strong	strong	+++
$S_{low}L\text{-}NoiseR$	weak	strong	++
$NoiseL\text{-}S_{low}R$	strong	weak	++

**Figure 7.** Summary of predicted effects due to cochlear efferents and cortical hemispheric preference in each of the four conditions with noise contralateral to filtered stimuli.

be lateralized, stronger in the right than the left ear (Khalfa et al., 1997; Philibert et al., 1998), and has accordingly been suggested to be part of the so called “right ear advantage” (Bilger et al., 1990; Kei et al., 1997; Khalfa et al., 1997; Newmark et al., 1997; Keogh et al., 2001). Noise is especially effective in eliciting this crossed efferent effect and, consequently, in the monaural speech condition with noise in the other ear, there is a systematic contralateral suppression of the neural response to the speech signal. Due to its asymmetry, the crossed efferent effect is expected to less strongly affect signal processing when it is delivered in the left ear. The efferent effect therefore interacts with the expected lateralized resonance between the filtered speech stimuli and the auditory cortical oscillations. Accordingly, we observed a stronger perceptual gain in the monaural  $S_{high}$  envelope condition than in the monaural  $S_{low}$  one (for prediction of cumulated effects, see Fig. 7).

The crossed cochlear efferent effect provides a plausible account for the observation that the dichotic effect in silence is abolished in binaural noise. When noise is present in both ears, crossed efferents have a positive net effect that results in selectively enhancing the neural response to the speech signal in a more pronounced way in the right ear. In our experiment, this effect is expected to enhance responses to weakly intelligible  $S_{high}$  speech in the preferred dichotic condition ( $S_{high}$  in right ear), thereby lowering global performance; and to enhance responses to the fairly intelligible  $S_{low}$  speech in the nonpreferred condition ( $S_{low}$  in right ear), thereby enhancing global performance. Accordingly, speech recognition in both dichotic conditions with background noise was intermediate between recognition in the two dichotic conditions in silence.

That the cochlear efferent system intervenes to counteract the hemispheric preference is plausible in our artificial experimental conditions. It is important to note, however, that by boosting auditory nerve processing more selectively in the right ear, the cochlear efferent system contributes to enhanced speech comprehension under normal stimulation conditions, generally noisy and with the full envelope spectrum reaching both ears.

#### Speech–left auditory cortex interaction at low gamma rate is behaviorally advantageous

The strongest argument in favor of asymmetric stimulus–brain resonance obtained from this series of experiments probably lies in the fact that enhanced neural activity could be observed in both auditory cortices in the preferred relative to nonpreferred dichotic condition despite weak (nonsignificant) mean behavioral effects. Furthermore, that neural activity correlated to performance in an asymmetric way underlines the relevance of stimulus–brain interactions on a gamma time scale to speech processing even



though these are not directly intelligible when presented alone. The contribution of this modulation range to speech processing remains to be clarified, as it is currently admitted that modulations beyond 10 Hz are marginally relevant to speech perception (Elliott and Theunissen, 2009). Yet voicing constants (30–70 ms), for instance, fall in the gamma range and gamma oscillations could play a key role in their perception. For instance, consonants with negative voice-onset time elicit one gamma wave on the EEG at voicing onset and another one at closure release (Trébuchon-Da Fonseca et al., 2005; Hoonhorst et al., 2009). However, the role of gamma cortical oscillations entrainment by speech stimuli is presumably more general than voicing encoding, and could lie in that it globally boosts auditory parsing at gamma rate. Automatic speech recognition devices classically use an initial step in which speech is chunked in 30 ms segments to perform subsequent computations, e.g., cepstral analysis (Noll, 1964). This chunking rate, chosen on mere engineering grounds, underlines the importance of gamma parsing as one of the first speech-specific computations performed in auditory cortex. That low gamma modulations in speech result in boosting gamma parsing is a tentative interpretation of the current data, which remains to be more directly addressed.

## References

- Abrams DA, Nicol T, Zecker S, Kraus N (2008) Right-hemisphere auditory cortex is dominant for coding syllable patterns in speech. *J Neurosci* 28:3958–3965.
- Besle J, Schevon CA, Mehta AD, Lakatos P, Goodman RR, McKhann GM, Emerson RG, Schroeder CE (2011) Tuning of the human neocortex to the temporal dynamics of attended events. *J Neurosci* 31:3176–3185.
- Bilger RC, Matthies ML, Hammel DR, Demorest ME (1990) Genetic implications of gender differences in the prevalence of spontaneous otoacoustic emissions. *J Speech Hear Res* 33:418–432.
- Boemio A, Fromm S, Braun A, Poeppel D (2005) Hierarchical and asymmetric temporal sensitivity in human auditory cortices. *Nat Neurosci* 8:389–395.
- Chait M, Greenberg S, Arai T, Simon J, Poeppel D (2005) Two time scales in speech processing. Paper presented at ISCA Workshop on Plasticity in Speech Perception, London, UK, June.
- Cohen J (1988) *Statistical power analysis for the behavior sciences*, 2nd edition. p 283. Hillsdale, NJ: Lawrence Erlbaum.
- Cooper NP, Guinan JJ Jr (2003) Separate mechanical processes underlie fast and slow effects of medial olivocochlear efferent activity. *J Physiol* 548:307–312.
- Drullman R, Festen JM, Plomp R (1994) Effect of temporal envelope smearing on speech reception. *J Acoust Soc Am* 95:1053–1064.
- Eickhoff SB, Stephan KE, Mohlberg H, Grefkes C, Fink GR, Amunts K, Zilles K (2005) A new SPM toolbox for combining probabilistic cytoarchitectonic maps and functional imaging data. *Neuroimage* 25:1325–1335.
- Eickhoff SB, Heim S, Zilles K, Amunts K (2006) Testing anatomically specified hypotheses in functional imaging using cytoarchitectonic maps. *Neuroimage* 32:570–582. Epub 2006 Jun 14.
- Elliott TM, Theunissen FE (2009) The modulation transfer function for speech intelligibility. *PLoS Comput Biol*. 5:e1000302.
- Giraud AL, Garnier S, Micheyl C, Lina G, Chays A, Chéry-Croze S (1997) Auditory efferents involved in speech-in-noise intelligibility. *Neuroreport* 8:1779–1783.
- Giraud AL, Kleinschmidt A, Poeppel D, Lund TE, Frackowiak RS, Laufs H (2007) Endogenous cortical rhythms determine cerebral specialization for speech perception and production. *Neuron* 56:1127–1134.
- Guinan JJ Jr, Gifford ML (1988) Effects of electrical stimulation of efferent olivocochlear neurons on cat auditory-nerve fibers. I. Rate-level functions. *Hear Res* 33:97–113.
- Hoonhorst I, Colin C, Markessis E, Radeau M, Deltenre P, Serniclaes W (2009) French native speakers in the making: from language-general to language-specific voicing boundaries. *J Exp Child Psychol* 104:353–366.
- Jäncke L, Shah NJ (2002) Does dichotic listening probe temporal lobe functions? *Neurology* 58:736–743.
- Jäncke L, Specht K, Shah JN, Hugdahl K (2003) Focused attention in a simple dichotic listening task: an fMRI experiment. *Brain Res Cogn Brain Res* 16:257–266.
- Kei J, McPherson B, Smyth V, Latham S, Loscher J (1997) Transient evoked otoacoustic emissions in infants: effects of gender, ear asymmetry and activity status. *Audiology* 36:61–71.
- Kell CA, Morillon B, Kounieher F, Giraud AL (2011) Lateralization of speech production starts in sensory cortices—a possible sensory origin of cerebral left dominance for speech. *Cereb Cortex* 21:932–937.
- Keogh T, Kei J, Driscoll C, Smyth V (2001) Distortion-product otoacoustic emissions in schoolchildren: effects of ear asymmetry, handedness, and gender. *J Am Acad Audiol* 12:506–513.
- Khalfa S, Morlet T, Micheyl C, Morgon A, Collet L (1997) Evidence of peripheral hearing asymmetry in humans: clinical implications. *Acta Otolaryngol* 117:192–196.
- Kumar UA, Vanaja CS (2004) Functioning of olivocochlear bundle and speech perception in noise. *Ear Hear* 25:142–146.
- Luo H, Poeppel D (2007) Phase patterns of neuronal responses reliably discriminate speech in human auditory cortex. *Neuron* 54:1001–1010.
- Micheyl C, Collet L (1996) Involvement of the olivocochlear bundle in the detection of tones in noise. *J Acoust Soc Am* 99:1604–1610.
- Morillon B, Lehongre K, Frackowiak RS, Ducorps A, Kleinschmidt A, Poeppel D, Giraud AL (2010) Neurophysiological origin of human brain asymmetry for speech and language. *Proc Natl Acad Sci U S A* 107:18688–18693.
- 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.
- Newmark M, Merlob P, Bresloff I, Olsha M, Attias J (1997) Click evoked otoacoustic emissions: inter-aural and gender differences in newborns. *J Basic Clin Physiol Pharmacol* 8:133–139.
- Noll AM (1964) Short-time spectrum and cepstrum techniques for vocal-pitch detection. *J Acoust Soc Am* 36:296–302.
- Obleser J, Eisner F, Kotz SA (2008) Bilateral speech comprehension reflects differential sensitivity to spectral and temporal features. *J Neurosci* 28:8116–8123.
- Oldfield RC (1971) The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia* 9:97–113.
- Philibert B, Veillet E, Collet L (1998) Functional asymmetries of crossed and uncrossed medial olivocochlear efferent pathways in humans. *Neurosci Lett* 253:99–102.
- Poeppel D (2003) The analysis of speech in different temporal integration windows: cerebral lateralization as ‘asymmetric sampling in time.’ *Speech Commun* 41:245–255.
- Schroeder CE, Lakatos P (2009) Low-frequency neuronal oscillations as instruments of sensory selection. *Trends Neurosci* 32:9–18.
- Shannon RV, Zeng FG, Kamath V, Wygonski J, Ekelid M (1995) Speech recognition with primarily temporal cues. *Science* 270:303–304.
- Silipo R, Greenberg S, Arai T (1999) Temporal constraints on speech intelligibility as deduced from exceedingly sparse spectral representations. In: *Proceedings of the 6th European Conference on Speech Communication and Technology (Eurospeech-99)*, pp 2687–2690.
- Telkemeyer S, Rossi S, Koch SP, Nierhaus T, Steinbrink J, Poeppel D, Obrig H, Wartenburger I (2009) Sensitivity of newborn auditory cortex to the temporal structure of sounds. *J Neurosci* 29:14726–14733.
- Trébuchon-Da Fonseca A, Giraud K, Badier JM, Chauvel P, Liégeois-Chauvel C (2005) Hemispheric lateralization of voice onset time (VOT) comparison between depth and scalp EEG recordings. *Neuroimage* 27:1–14.
- Van Tasell DJ, Soli SD, Kirby VM, Widin GP (1987) Speech waveform envelope cues for consonant recognition. *J Acoust Soc Am* 82:1152–1161.
- Westerhausen R, Hugdahl K (2008) The corpus callosum in dichotic listening studies of hemispheric asymmetry: a review of clinical and experimental evidence. *Neurosci Biobehav Rev* 32:1044–1054.
- Zaehle T, Wüstenberg T, Meyer M, Jäncke L (2004) Evidence for rapid auditory perception as the foundation of speech processing: a sparse temporal sampling fMRI study. *Eur J Neurosci* 20:2447–2456.
- Zatorre RJ, Belin P (2001) Spectral and temporal processing in human auditory cortex. *Cereb Cortex* 11:946–953.
- Zatorre RJ, Gandour JT (2008) Neural specializations for speech and pitch: moving beyond the dichotomies. *Philos Trans R Soc Lond B Biol Sci* 363:1087–1104.