Published in final edited form as:
Open J Neurosci. ; 3: .

Silent lipreading and covert speech production suppress processing of non-linguistic sounds in auditory cortex

Marja H. Balk[1,2], **Heini Kari**[1], **Jaakko Kauramäki**[1], **Jyrki Ahveninen**[3], **Mikko Sams**[1], **Taina Autti**[2], and **Iiro P. Jääskeläinen**[1,3,4]

[1]Brain and Mind Laboratory, Department of Biomedical Engineering and Computational Science (BECS), Aalto University School of Science, Espoo, Finland [2]Department of Radiology, HUS Medical Imaging Center, Helsinki University Central Hospital and University of Helsinki, Finland [3]Harvard Medical School – Athinoula A. Martinos Center for Biomedical Imaging, Department of Radiology, Massachusetts General Hospital, Charlestown, MA, US [4]Advanced Magnetic Imaging Center, Aalto University School of Science, Espoo, Finland

Abstract

Previous studies have suggested that speech motor system mediates suppression by silent lipreading of electromagnetic auditory cortex responses to pure tones at about 100 ms from sound onset. We used sparse sampling functional magnetic resonance imaging (fMRI) at 3 Tesla to map auditory-cortex foci of suppressant effects during silent lipreading and covert self-production. Streams of video clips were presented simultaneously with 1/3 octave noise bursts centered at 250 Hz (low frequency, LF) or 2000 Hz (mid-frequency, MF), or during no auditory stimulation. In different conditions, the subjects were a) to press a button whenever they lipread the face articulate the same consecutive Finnish vowels /a/, /i/, /o/, and /y/, b) covertly selfproducing vowels while viewing still face image, or c) to press a button whenever a circle pictured on top of the lips expanded into oval shape of the same orientation twice in a row. The regions of interest (ROIs) within the superior temporal lobes of each hemisphere were defined by contrasting MF and LF stimulation against silence. Contrasting the nonlinguistic (*i.e.*, expanding circle) *vs.* linguistic (*i.e.*, lipreading and covert self-production) conditions within these ROIs showed significant suppression of hemodynamic activity to MF sounds in the linguistic condition in left hemisphere first transverse sulcus (FTS) and right hemisphere superior temporal gyrus (STG) lateral to Heschl's sulcus (HS). These findings suggest that the speech motor system mediates suppression of auditory-cortex processing of non-linguistic sounds during silent lipreading and covert self-production in left hemisphere FST and right hemisphere STG lateral to HS.

Keywords

auditory cortex; lipreading; covert speech production; speech motor theory; functional magnetic resonance imaging; tonotopy

Corresponding Author & Address: ***Marja H. Balk***, Department of Radiology, HUS Medical Imaging Center, Helsinki University Central Hospital, P.O. BOX 180, FIN-00029 HUS, Finland; marja.balk@hus.fi; Tel: +358 50427 2843; Fax +358 9 472 3182, ***Iiro P. Jääskeläinen***, Brain and Mind Laboratory, Department of Biomedical Engineering and Computational Science, Aalto University, P.O. Box 12200, FIN-00076 Aalto University, Finland; iiro.jaaskelainen@aalto.fi; Tel: +358 50560 9503; Fax: +358 9 470 23182.

CONFLICTS OF INTEREST

The authors declare no conflicts of interest.

INTRODUCTION

Audiovisual integration has a fundamental role in human information processing. During development, one is exposed to highly correlated auditory and visual speech stimulation [1, 2]. Indeed, multisensory perception is essential in every-day communication; audiovisual interactions often accelerate and improve perception. For example, it has been shown that seeing the speaker's articulatory lip movements enhance speech comprehension especially when there is acoustic noise [3]. In contrast, conflicting audiovisual features can degrade or even alter perception. In the so-called McGurk illusion conflicting visual and auditory phonemes can produce illusory percept of a third phoneme (*e.g.*, visual /ga/ combined to auditory /ba/ is perceived as /da/, see [4]).

Recent functional magnetic resonance imaging (fMRI) studies have provided findings on the neural basis of how seeing articulatory gestures influences auditory cortex speech processing. Audiovisual speech has been frequently shown to elicit hemodynamic activity in the posterior superior temporal sulcus (STS) /superior temporal gyrus (STG) [5–11] and it has been shown that seeing articulatory gestures can modulate even primary auditory cortex hemodynamic activity [12, 13]. Further, attention to audiovisual stimuli has been observed to enhance hemodynamic responses in the planum temporale (PT, a part of the secondary auditory sensory cortex [14]). These findings on audiovisual speech processing somewhat deviate from findings (initially suggested by primate electrophysiological studies [15]), showing that human auditory cortex is organized to process auditory stimuli in parallel anterior "*what*" and posterior "*where*" processing pathways with speech being processed within the "*anterior*" stream [16, 17]. Supporting this notion, previous neuroimaging studies on auditory speech have consistently demonstrated hemodynamic activations in the areas anterior to Heschl's gyrus (HG, *i.e.* primary auditory cortex [18–23]), even bilaterally [17, 22, 24].

One possible explanation for the apparent discrepancy in findings between studies on the neural basis of auditory *vs.* audiovisual speech processing is that the dorsal "*where*" pathway involves a broader set of functions than mere spatial processing, specifically mapping of auditory inputs to motor schemas (*i.e.*, the "*how*" pathway) [25]. Furthermore, the effects of lipreading are not restricted to the sensory specific areas (*i.e.* visual and auditory cortices), but have been reported to activate the Broca's area [9, 10, 26], motor cortex [10, 26, 27], posterior parietal cortex [9, 26–28], claustrum [8], and insular cortex [9]. Such large scale of motor system activations tentatively suggest that audiovisual speech perception is closely related to speech production and it has been specifically suggested that seeing speech modulates auditory cortex *via* top-down inputs from the speech motor system [2, 10, 25, 29].

Magnetoencephalography (MEG) studies have supported the notion that the speech-motor system mediates auditory-cortex modulation during lipreading. Suppression of electromagnetic N1 response to audiovisual phonetic stimuli, generated ~100 ms from sound onset in the auditory cortex and its posterior areas, has been reported as compared with responses to auditory phonetic stimuli [2, 30–32]. Furthermore, silent lipreading suppresses N1 responses to phonemic F2 transitions [33] and suppression of N1 to auditory speech stimuli has been observed during overt and covert speech production [34]. In our recent study, we observed that N1 responses to pure tones were suppressed similarly during lipreading and covert speech production most prominently in the left hemisphere. These effects were seen for all tone frequencies ranging from 125 to 8000 Hz, but due the ill-posed nature of the MEG inverse estimation the precise anatomical loci of this effect remained ambiguous [35].

There are at least two auditory cortex source generators that contribute to the N1 response [30, 36, 37] and, further, fMRI studies have demonstrated several tonotopic maps in the human superior temporal lobe [38–43]. While the functional specialization of these tonotopically organized areas have not been elucidated, the loci of these areas is such that they could potentially contribute to the generation of the N1 response [44]. Thus, it is feasible to assume that seeing articulatory gestures might suppress activity in some of these tonotopic areas and result in suppression of responses to even non-linguistic auditory stimuli.

Here, we studied with fMRI how the human auditory cortex processes narrow-band (1/3 octave) noise bursts centered at low frequencies (*i.e.*, out of frequency band that is critical for speech processing) and mid-frequencies (*i.e.*, within the frequency band that is critical for speech processing) during presentation of linguistic and non-linguistic visual stimuli. We hypothesized that seeing articulatory gestures suppresses auditory-cortex processing of the narrow-band noise bursts compared to the control non-linguistic condition. We restricted the region of interest to the superior temporal plane covering the primary and secondary auditory cortex and further set forth to investigate whether the suppressive effects specifically concern some of the tonotopic areas as described by Talavage *et al.* 2004 [40].

MATERIALS AND METHODOLOGY

Subjects

Fifteen healthy right-handed native Finnish speakers (four men; age range 21 – 55; average = 26.9 years) gave an informed consent to participate in the experiment. Two subjects were discarded due technical problems during fMRI scanning. The subjects had normal hearing and normal or corrected-to-normal vision, and they reported having no history of neurological or auditory diseases or symptoms. The study protocol was approved by the Ethics Committee of Hospital District of Helsinki and Uusimaa, Finland and the research was conducted in accordance with the Helsinki declaration. Subjects received no financial compensation for their participation.

Stimuli and Task

The auditory stimuli were 100 ms noise bursts, which had a bandwidth of 1/3 octave and a center frequency equal to either 250 Hz (low frequency, LF) or 2000 Hz (mid-frequency, MF; Figure 1). The frequency borders for the LF noise burst were 223 and 281 Hz and those for the MF noise burst were 1782 and 2245 Hz. Both sound clips had 5-ms Hanning-windowed onset and offset rams. The onset-to-onset inter-stimulus interval was 500 ms. Each condition (LF, MF, and silence) was presented for 30 s in random order. The sound files were generated with Matlab (R14*,* MathWorks, Natlick, MA, USA). We used 44.1 kHz sampling rate with 16 bit precision.

The visual stimuli were the same as the ones used in our previous experiment [35]. In the lipreading condition a female face articulated Finnish vowels /a/, /i/, /o/, and /y/. The digitized videoclip of each articulation lasted for 1320, 1360, 1400, and 1440 ms, respectively. We combined the videoclips of the articulations in a pseudorandom order to create a continuous 10-minute video. The subjects were requested to press a magnet-compatible button whenever they detected two consecutive same-category articulations. In the expanding-circle condition a blue transparent circle was overlaid on the mouth area of the still-face image and the circle transformed into an oval of four alternative directions (horizontal, vertical, right oblique, and left oblique). The time scale and spatial frequency of the circle transformation into an oval approximated that of the mouth opening during the lipreading condition. The videoclips of the circle transformations to ovals were concatenated pseudorandomly in the same way as the vowel videoclips in the lipreading condition to form a continuous video of 10 min. The subjects were instructed to press the response paddle whenever they detected occurrence of two consecutive oval expression of the same direction. In covert-production of vowels condition we used still face of the same female and the subjects were asked to covertly produce vowels at roughly the visual stimulus presentation rate of the other conditions. The screen resolution was 640 × 480 pixels at 60 Hz with 32- bit color depth.

Each visual condition (lipreading, expanding-circle, and covert self-production) formed one 10-min run, in which we used a block design with alternating 30-s noise bursts (LF, MF) and silence blocks (Figure 1). Each run consisted of 20 counterbalanced auditory blocks. The order of the runs was randomized across subjects. During lipreading and the expanding-circle conditions the target of the one-back task occurred 10 times per each run. Each visual condition was repeated twice in random order, thus the total functional scanning time was 60 minutes.

Prescanning Procedures

Before the experiment, the tasks were explained and the subjects were allowed to practice one test run per condition to familiarize them with the paradigm. The stimuli in the practice tasks were different from the tasks used in the main experiment.

Prior the measurements the individual hearing threshold was assessed using LF noise burst as a test sound and the intensity of the test sound was adjusted 40 dB above the hearing threshold. As in the fMRI environment the use of pneumatic headphones might induce sound attenuation, especially in high frequencies, the LF and MF noise bursts were individually adjusted to match the same hearing level. In the attenuation test the subjects heard consecutively LF and MF noise bursts (= attenuation pair) and they had to choose the loudest of the attenuation pair using a response paddle. The response automatically changed the attenuation of the MF noise burst for the next attenuation pair and the assessments of attenuation pairs were continued until the equilibrium of the attenuation was achieved. During both adjustments the subjects were in the MRI bore.

Data Acquisition

We scanned the subjects with 3 Tesla GE Signa MRI (Milwaukee, WI, USA) with eight-channel quadrature head coil. We used sparse sampling technique [45] in which the functional imaging of 0.8 s was followed by the 9.2 s of silence. The coolant pump of the magnet was switched off during prescanning adjustments of the sound and during functional imaging to further reduce acoustic noise. The functional gradient-echo echo-planar (GE-EP) T2*-weighted volumes had the following parameters: repetition time 800 ms, echo time 30 ms, matrix 64 × 64, flip angle 90 degrees, field of view 22 cm, slice thickness 3 mm, no gap, 12 near-axial orientated slices, delay between the acquisition of the GE-EP volumes was 9.2 s. The effective voxel size was 3.43 × 3.43 × 3.00 mm. The functional volume consisted of 12 slices set parallel to the superior temporal sulcus – with the most superior slices covering the HG. Each experimental run produced 60 GE-EP volumes. The subjects were instructed to avoid body movements throughout the experiment with the head fixed with foam cushions on both sides.

After functional imaging the whole-head GE-EP images were obtained in the same slice orientation as the functional volume, followed by the 3D T1 weighted axial slices for co-registration and T2-Fluid Attenuation Inversion Recovery (FLAIR) images. All the auditory stimuli were presented binaurally to the participants with MRI-fitted narrow tubes that had ear plugs with pores in the center attached to them. The visual stimuli were projected *via* a mirror-system stationed onto the head coil inside the magnet bore. Subjects were instructed to focus to the mouth region of the face. We delivered the stimuli with the Presentation (Neurobehavioral System, Albany, CA, USA) program.

Data Analysis

The data analysis was conducted with BrainVoyager QX software version 2.1 (Brain Innovation, Maastricht, The Netherlands [46]). Two participants were discarded prior to analysis and one participant's one run was discarded because of technical reasons. The three volumes of each run were discarded to allow for T1-saturation, thus leaving 342 GE-EP - volumes of each subject into final analysis. Preprocessing for single-participant data included 3D motion correction, Gaussian spatial smoothing (full-width-half-maximum 5 mm), and linear trend removal. Slice timing correction was not included as we used the sparse sampling technique. Each individual’s functional images were co-registered with their high-resolution anatomical images and transformed into the standard Talairach coordinate system.

The group data were analyzed with multi-subject random-effects general linear model (GLM) in standard space. Statistical analysis was carried out using GLM with three auditory stimuli (LF noise burst, MF noise burst, and silence baseline) as independent predictors. The model was not convolved to a hemodynamic response function because of the sparseness of the data.

First we created region of interests within which we carried out the further analyses. Using a GLM, LF noise burst and MF noise burst stimuli were contrasted as separate predictors against silence in each condition at the significance level $P < 0.05$. The ensuing clusters encompassing the auditory cortical areas in each hemisphere in each condition were merged to form one ROI per each hemisphere (Figure 2). The further statistical group-level contrasts were carried out inside these ROIs using significance level $P < 0.05$.

The high-resolution images were brought into anterior commisura (AC) – posterior commisura (PC) space. Further segmentation according to their grey and white matter and hemispheric segregation was done with the BrainVoyager semiautomatic segmentation program using sigma filtering and automatic bridge removal algorithms. As signal intensity homogeneity varied extensively in the temporal and frontal lobes all anatomical images needed additional manual correction. Reconstruction of the cortex allowed us to display activations on the gyri and sulci in both flattened and inflated brain views: the group analysis was presented on the inflated or flattened hemispheres of a single subject’s brain.

Analysis of Behavioral Task

We measured both reaction time (RT) for detecting the target stimuli and the hit rate. The RT was measured from the beginning of the visual target clip to the button press, thus increasing the RT compared to simple target identification as the video clip type could be identified at earliest 300–400 ms from onset (See Figure 1A). The hit rate (HR) was the proportion of targets detected. We compared hit rate and RT between the different conditions using Student’s t-test.

RESULTS AND OBSERVATIONS

Behavioral Data

There were no significant differences between HRs to targets between the expanding-circle (mean ± SD = 82.7 ± 3.4%) and the lipreading (76.7 ± 4.5%) conditions ($p > 0.14$). The RT was 1512 ± 26 ms for detection of target vowel articulations and 1418 ± 23 ms for detection of expanding oval targets. RT for the expanding circle targets was significantly faster than that for vowel articulation detection ($p < 0.05$); as the oval direction was apparent from the first deformation compared to mouth movements, where the distinct differences appeared only later.

Imaging Data

Conjunction analysis of all group-level activations disclosed a wide range of temporal-lobe activations that extended to subcortical regions (Figure 2). The temporal-lobe cortical areas revealed by this analysis served as the region-of-interest (ROI) for further planned contrasts between the different experimental conditions.

Figure 3 depicts significant group activations of MF noise bursts and LF noise bursts in the flattened right and left superior temporal ROIs of one subject. The sounds activated the middle and posterior STS, including Heschl's sulcus (HS) and HG, with some extensions into STG bilaterally. In all conditions the activations caused by the MF and LF noise bursts were more extended in the left than in the right hemisphere.

Activity elicited by the MF noise bursts encompassed a slightly more limited region in the superior temporal plane than the hemodynamic responses to LF noise bursts; the activations caused by the LF noise bursts surrounded the activations caused by the MF noise bursts. The activations were more anterior in the right than in the left hemisphere, and there was a slight left-right asymmetry in the posterior STS activations in all conditions.

When contrasting the non-linguistic (*i.e.*, expanding circle) *vs.* linguistic (*i.e.*, lipreading and covert self-production) conditions within the temporal lobe ROIs, significant suppression of hemodynamic activity to MF noise bursts was observed in the linguistic condition in the left hemisphere first transverse sulcus (FTS) and right hemisphere STG lateral to HS (Figure 4). There were no significant differences in the contrast between the lipreading and covert self-production conditions.

DISCUSSION

In the present study, we investigated using fMRI in which auditory cortical regions silent lipreading and covert self-production suppresses processing of simple auditory stimuli. By examining with two different frequency bands we found that contrasting non-linguistic (*i.e.*, expanding circle) *vs.* linguistic (*i.e.*, lipreading and covert self-production) conditions during MF noise burst stimulation results in bilateral, but asymmetric activations on the superior temporal cortex involving FTS in the left and STG lateral to HS in the right hemisphere (Figure 4). This suggests that speech motor system modulates processing of non-linguistic sounds at speech-relevant frequencies by suppressing hemodynamic reactivity of specific auditory cortical areas. Previously, several tonotopically organized areas have been documented in auditory cortex suggesting that the auditory cortex is functionally highly heterogeneous. We tentatively propose that the suppressed areas in the left FTS and in the right STG could be related to tonotopic areas described by Talavage *et al.* [40].

The activations of the LF and MF noise bursts overlapped the auditory cortex in the left and right hemispheres within the auditory cortical regions of interest across the conditions (Figure 3). Since we specifically selected LF and MF noise bursts in the present study, the results are not easily comparable to the previous tonotopic mapping results, *per se* [38–42]. However, it can be speculated that the areas activated by MF noise bursts would be more relevant for speech processing than those activated by LF sounds due to the MF sounds occupying the frequency band that is critical for speech perception. While there seem to be some frequencydependent differences across the conditions as shown in Figure 3, it has to be noted that the maps were threshold against the silent baseline rather than contrasted between the conditions. Nevertheless, the activations in the right hemisphere disclosed an area more anteriorly than in the left hemisphere. Such activity distributions could reflect hemispheric differences in processing auditory stimuli: right hemisphere has been suggested to process acoustic sound features such as pitch [47] and the left hemisphere has been suggested to specialized in processing speech-related temporal dynamics [18, 23, 48]. Noting that even non-semantic non-speech audiovisual information has access to superior temporal cortex [49–51], the vast hemodynamic activations in our study could also be associated to audiovisual integration. All the conditions showed left-hemispheric activations posterior to HG and medial to planum temporale or adjacent to somatosensory areas of tongue and pharynx (Figure 3). This could tentatively suggest that speech-related stimuli have access to speech motor areas and thus, support the left-hemispheric lateralization to linguistic information or even mirroring of action into perception [27]. However, we cannot exclude subvocalization during the covert speech production, although we encouraged the subjects to avoid mouth movements and to articulate the vowels only in their minds in the covert articulation condition. Because the ROIs were limited to auditory cortex no visual cortex or other heteromodal cortices were imaged in the present study.

Contrasting the non-linguistic (*i.e.*, expanding circle) conditions to linguistic (*i.e.,* lipreading and covert self-production) revealed asymmetric supratemporal activations (Figure 4). Within the temporal lobe ROIs significant suppression of hemodynamic activity to speech-related MF noise bursts was shown in the linguistic conditions in the left hemisphere FTS and right hemisphere STG lateral to HS. In previous studies, the auditory cortex has been documented to contain several tonotopic areas with the functions, interactions, and connectivity of these areas presumably highly heterogeneous. In the tonotopic mapping experiment Talavage *et al.* revealed several tonotopically organized areas using amplitude-modulated noise [40]. Our FTS activation in the left hemisphere could correlate with Talavage’s “higher-frequency sensitivity endpoint 2“. Interpretation of the right hemisphere activation is more speculative. Although our right hemispheric STG activation could be seen as corresponding to Talavage’s “higher-frequency endpoint 5”, one has to be cautious as Talavage’s tonotopic mapping involved only the left hemisphere. Striem-Amit *et al.* investigated bilateral tonotopic mapping across left and right supratemporal lobes with rising tone chirps ranging from 250 to 4,000 Hz [43]. Our asymmetric suppressant activations to speech-related MF noise bursts could be pinpointed to the areas where Striem-Amit *et al.* showed activations at medium frequencies. We failed to see any significant differences in the contrast between the lipreading and covert self-production conditions that tentatively suggests that there might have been similar mechanisms at work.

Our present study restrained the analysis on the supratemporal lobes, as our approach was merely on the primary auditory cortex and the surrounding areas. It is then plausible that our region specific analysis could have excluded other speech related regions that are known to support lipreading such as Broca´s area [52]. Likewise, our slice- and ROI selection excluded visual cortex or other heteromodal cortices [53]. The limited temporal resolution of fMRI is a further limitation of our experiment. Therefore, the speech-related hemodynamic activations cannot be related to suppressed N1 observed in our previous MEG experiment [35] in a straightforward manner. Nonetheless, with similar experimental setup MEG demonstrated N1 response suppression in the superior temporal lobes during lipreading and covert speech-production and fMRI exhibited suppressant activations in the left FTS and in the right STG lateral to HS. Together, the present results combined with previous MEG study [35] show that the speech-related suppression was bilateral, but with distinct areas suppressed asymmetrically. In both studies, suppression effect was highly similar between covert speech self-production and lipreading tasks, thus suggesting similar underlying mechanism that mediates suppression of auditory processing during lipreading and covert self-production. Importantly, the present results suggest that the top-down input from speech motor areas induced suppression lasting several seconds (besides sub-second and transient suppression revealed by MEG) in at least two distinct areas the auditory cortex, more prominently in the left hemisphere.

CONCLUSION

In conclusion, in the present study lipreading and covert speech self-production suppressed processing of non-linguistic sounds in the auditory cortex. We suggest that speech-related suppressant effect arises in tonotopic subareas of auditory cortex, in the left FTS and in the right STG lateral to HS.

Acknowledgments

This study was financially supported by the Academy of Finland, the EVO Fund of the HUS Medical Imaging Center, Helsinki University Central Hospital, and by the US National Institutes of Health grants R21DC010060, R01MH083744, R01HD040712, and R01NS037462. We thank all the volunteers for participating.

REFERENCES

1. Lewkowicz DJ. Infant perception of audio-visual speech synchrony. Dev Psychol. 2010; 46:66–77. [PubMed: 20053007]
2. Jaaskelainen IP. The role of speech production system in audiovisual speech perception. Open Neuroimaging Journal. 2010; 4:30–36. [PubMed: 20922046]
3. Sumby WH, Pollack I. Visual contribution to speech intelligibility in noise. J Acoust Soc America. 1954; 26:212–215.
4. McGurk H, MacDonald J. Hearing lips and seeing voices. Nature. 1976; 264:746–748. [PubMed: 1012311]
5. Beauchamp MS, Argall BD, Bodurka J, Duyn JH, Martin A. Unraveling multisensory integration: patchy organization within human STS multisensory cortex. Nat Neurosci. 2004; 7:1190–1192. [PubMed: 15475952]
6. Campbell R, MacSweeney M, Surguladze S, Calvert G, McGuire P, Suckling J, Brammer MJ, David AS. Cortical substrates for the perception of face actions: an fMRI study of the specificity of activation for seen speech and for meaningless lower-face acts (gurning). Brain Res Cogn Brain Res. 2001; 12:233–243. [PubMed: 11587893]
7. Murase M, Saito DN, Kochiyama T, Tanabe HC, Tanaka S, Harada T, Aramaki Y, Honda M, Sadato N. Cross-modal integration during vowel identification in audiovisual speech: A functional magnetic resonance imaging study. Neurosci Lett. 2008; 434:71–76. [PubMed: 18280656]
8. Olson IR, Gatenby JC, Gore JC. A comparison of bound and unbound audio–visual information processing in the human cerebral cortex. Cogn Brain Res. 2002; 14:129–138.
9. Miller LM, D'Esposito M. Perceptual fusion and stimulus coincidence in the cross-modal integration of speech. J Neurosci. 2005; 25:5884–5893. [PubMed: 15976077]
10. Skipper JI, van Wassenhove V, Nusbaum HC, Small SL. Hearing lips and seeing voices: how cortical areas supporting speech production mediate audiovisual speech perception. Cereb Cortex. 2007; 17:2387–2399. [PubMed: 17218482]
11. Wright TM, Pelphrey KA, Allison T, McKeown MJ, McCarthy G. Polysensory interactions along lateral temporal regions evoked by audiovisual speech. Cereb Cortex. 2003; 13:1034–1043. [PubMed: 12967920]
12. Calvert GA, Bullmore ET, Brammer MJ, Campbell R, Williams SC, McGuire PK, Woodruff PW, Iversen SD, David AS. Activation of auditory cortex during silent lipreading. Science. 1997; 276:593–596. [PubMed: 9110978]
13. Pekkola J, Ojanen V, Autti T, Jaaskelainen IP, Mottonen R, Tarkiainen A, Sams M. Primary auditory cortex activation by visual speech: an fMRI study at 3 T. Neuroreport. 2005; 16:125–128. [PubMed: 15671860]

14. Pekkola J, Ojanen V, Autti T, Jaaskelainen IP, Mottonen R, Sams M. Attention to visual speech gestures enhances hemodynamic activity in the left planum temporale. Hum Brain Mapp. 2006; 27:471–477. [PubMed: 16161166]
15. Tian B, Reser D, Durham A, Kustov A, Rauschecker JP. Functional specialization in rhesus monkey auditory cortex. Science. 2001; 292:290–293. [PubMed: 11303104]
16. Alain C, Arnott SR, Hevenor S, Graham S, Grady CL. “What” and"where” in the human auditory system. Proc Natl Acad Sci U S A. 2001; 98:12301–12306. [PubMed: 11572938]
17. Ahveninen J, Jaaskelainen IP, Raij T, Bonmassar G, Devore S, Hamalainen M, Levanen S, Lin FH, Sams M, Shinn-Cunningham BG, Witzel T, Belliveau JW. Task-modulated "what" and "where" pathways in human auditory cortex. Proc Natl Acad Sci U S A. 2006; 103:14608–14613. [PubMed: 16983092]
18. Binder JR, Frost JA, Hammeke TA, Bellgowan PS, Springer JA, Kaufman JN, Possing ET. Human temporal lobe activation by speech and nonspeech sounds. Cereb Cortex. 2000; 10:512–528. [PubMed: 10847601]
19. Binder JR, Liebenthal E, Possing ET, Medler DA, Ward BD. Neural correlates of sensory and decision processes in auditory object identification. Nat Neurosci. 2004; 7:295–301. [PubMed: 14966525]
20. Liebenthal E, Binder JR, Spitzer SM, Possing ET, Medler DA. Neural substrates of phonemic perception. Cereb Cortex. 2005; 15:1621–1631. [PubMed: 15703256]
21. Narain C, Scott SK, Wise RJ, Rosen S, Leff A, Iversen SD, Matthews PM. Defining a left-lateralized response specific to intelligible speech using fMRI. Cereb Cortex. 2003; 13:1362–1368. [PubMed: 14615301]
22. Obleser J, Boecker H, Drzezga A, Haslinger B, Hennenlotter A, Roettinger M, Eulitz C, Rauschecker JP. Vowel sound extraction in anterior superior temporal cortex. Hum Brain Mapp. 2006; 27:562–571. [PubMed: 16281283]
23. Scott SK, Blank CC, Rosen S, Wise RJ. Identification of a pathway for intelligible speech in the left temporal lobe. Brain. 2000; 123:2400–2406. [PubMed: 11099443]
24. DeWitt I, Rauschecker JP. Phoneme and word recognition in the auditory ventral stream. Proc Natl Acad Sci U S A. 2012; 109:E505–E514. [PubMed: 22308358]
25. Hickok G, Poeppel D. Towards a functional neuroanatomy of speech perception. Trends Cogn Sci. 2000; 4:131–138. [PubMed: 10740277]
26. Ojanen V, Mottonen R, Pekkola J, Jaaskelainen IP, Joensuu R, Autti T, Sams M. Processing of audiovisual speech in Broca's area. Neuroimage. 2005; 25:333–338. [PubMed: 15784412]
27. Nishitani N, Hari R. Viewing lip forms: cortical dynamics. Neuron. 2002; 36:1211–1220. [PubMed: 12495633]
28. Bernstein LE, Auer ET Jr, Wagner M, Ponton CW. Spatiotemporal dynamics of audiovisual speech processing. Neuroimage. 2008; 39:423–435. [PubMed: 17920933]
29. Hickok G, Poeppel D. The cortical organization of speech processing. Nat Rev Neurosci. 2007; 8:393–402. [PubMed: 17431404]
30. Jaaskelainen IP, Ojanen V, Ahveninen J, Auranen T, Levanen S, Mottonen R, Tarnanen I, Sams M. Adaptation of neuromagnetic N1 responses to phonetic stimuli by visual speech in humans. Neuroreport. 2004; 15:2741–2744. [PubMed: 15597045]
31. Klucharev V, Mottonen R, Sams M. Electrophysiological indicators of phonetic and non-phonetic multisensory interactions during audiovisual speech perception. Brain Res Cogn Brain Res. 2003; 18:65–75. [PubMed: 14659498]
32. van Wassenhove V, Grant KW, Poeppel D. Visual speech speeds up the neural processing of auditory speech. Proc Natl Acad Sci U S A. 2005; 102:1181–1186. [PubMed: 15647358]
33. Jaaskelainen IP, Kauramaki J, Tujunen J, Sams M. Formant transition-specific adaptation by lipreading of left auditory cortex N1m. Neuroreport. 2008; 19:93–97. [PubMed: 18281900]
34. Numminen J, Curio G. Differential effects of overt, covert and replayed speech on vowel-evoked responses of the human auditory cortex. Neurosci Lett. 1999; 272:29–32. [PubMed: 10507535]
35. Kauramaki J, Jaaskelainen IP, Hari R, Mottonen R, Rauschecker JP, Sams M. Lipreading and covert speech production similarly modulate human auditory-cortex responses to pure tones. J Neurosci. 2010; 30:1314–1321. [PubMed: 20107058]

36. Sams M, Hari R, Rif J, Knuutila J. The Human Auditory Sensory Memory Trace Persists about 10 sec: Neuromagnetic Evidence. J Cogn Neurosci. 1993; 5:363–370.

37. Woods DL. The component structure of the N1 wave of the human auditory evoked potential. Electroencephalogr Clin Neurophysiol Suppl. 1995; 44:102–109. [PubMed: 7649012]

38. Formisano E, Kim DS, Di Salle F, van de Moortele PF, Ugurbil K, Goebel R. Mirror-symmetric tonotopic maps in human primary auditory cortex. Neuron. 2003; 40:859–869. [PubMed: 14622588]

39. Talavage TM, Ledden PJ, Benson RR, Rosen BR, Melcher JR. Frequency-dependent responses exhibited by multiple regions in human auditory cortex. Hear Res. 2000; 150:225–244. [PubMed: 11077206]

40. Talavage TM, Sereno MI, Melcher JR, Ledden PJ, Rosen BR, Dale AM. Tonotopic organization in human auditory cortex revealed by progressions of frequency sensitivity. J Neurophysiol. 2004; 91:1282–1296. [PubMed: 14614108]

41. Schonwiesner M, von Cramon DY, Rubsamen R. Is it tonotopy after all? Neuroimage. 2002; 17:1144–1161. [PubMed: 12414256]

42. Langers DR, Backes WH, van Dijk P. Representation of lateralization and tonotopy in primary versus secondary human auditory cortex. Neuroimage. 2007; 34:264–273. [PubMed: 17049275]

43. Striem-Amit E, Hertz U, Amedi A. Extensive cochleotopic mapping of human auditory cortical fields obtained with phase-encoding FMRI. PloS One. 2011; 6:e17832. [PubMed: 21448274]

44. Jaaskelainen IP, Ahveninen J, Bonmassar G, Dale AM, Ilmoniemi RJ, Levanen S, Lin FH, May P, Melcher J, Stufflebeam S, Tiitinen H, Belliveau JW. Human posterior auditory cortex gates novel sounds to consciousness. Proc Natl Acad Sci U S A. 2004; 101:6809–6814. [PubMed: 15096618]

45. Hall DA, Haggard MP, Akeroyd MA, Palmer AR, Summerfield AQ, Elliott MR, Gurney EM, Bowtell RW. "Sparse" temporal sampling in auditory fMRI. Hum Brain Mapp. 1999; 7:213–223. [PubMed: 10194620]

46. Goebel R, Esposito F, Formisano E. Analysis of functional image analysis contest (FIAC) data with brainvoyager QX: From single-subject to cortically aligned group general linear model analysis and self-organizing group independent component analysis. Hum Brain Mapp. 2006; 27:392–401. [PubMed: 16596654]

47. Zatorre RJ, Belin P. Spectral and temporal processing in human auditory cortex. Cereb Cortex. 2001; 11:946–953. [PubMed: 11549617]

48. Belin P, Zatorre RJ, Ahad P. Human temporal-lobe response to vocal sounds. Cogn Brain Res. 2002; 13:17–26.

49. Calvert GA, Hansen PC, Iversen SD, Brammer MJ. Detection of audio-visual integration sites in humans by application of electrophysiological criteria to the BOLD effect. Neuroimage. 2001; 14:427–438. [PubMed: 11467916]

50. Dhamala M, Assisi CG, Jirsa VK, Steinberg FL, Scott Kelso JA. Multisensory integration for timing engages different brain networks. Neuroimage. 2007; 34:764–773. [PubMed: 17098445]

51. Noesselt T, Rieger JW, Schoenfeld MA, Kanowski M, Hinrichs H, Heinze HJ, Driver J. Audiovisual temporal correspondence modulates human multisensory superior temporal sulcus plus primary sensory cortices. J Neurosci. 2007; 27:11431–11441. [PubMed: 17942738]

52. Paulesu E, Perani D, Blasi V, Silani G, Borghese NA, De Giovanni U, Sensolo S, Fazio F. A functional-anatomical model for lipreading. J Neurophysiol. 2003; 90:2005–2013. [PubMed: 12750414]

53. Macaluso E, George N, Dolan R, Spence C, Driver J. Spatial and temporal factors during processing of audiovisual speech: a PET study. Neuroimage. 2004; 21:725–732. [PubMed: 14980575]

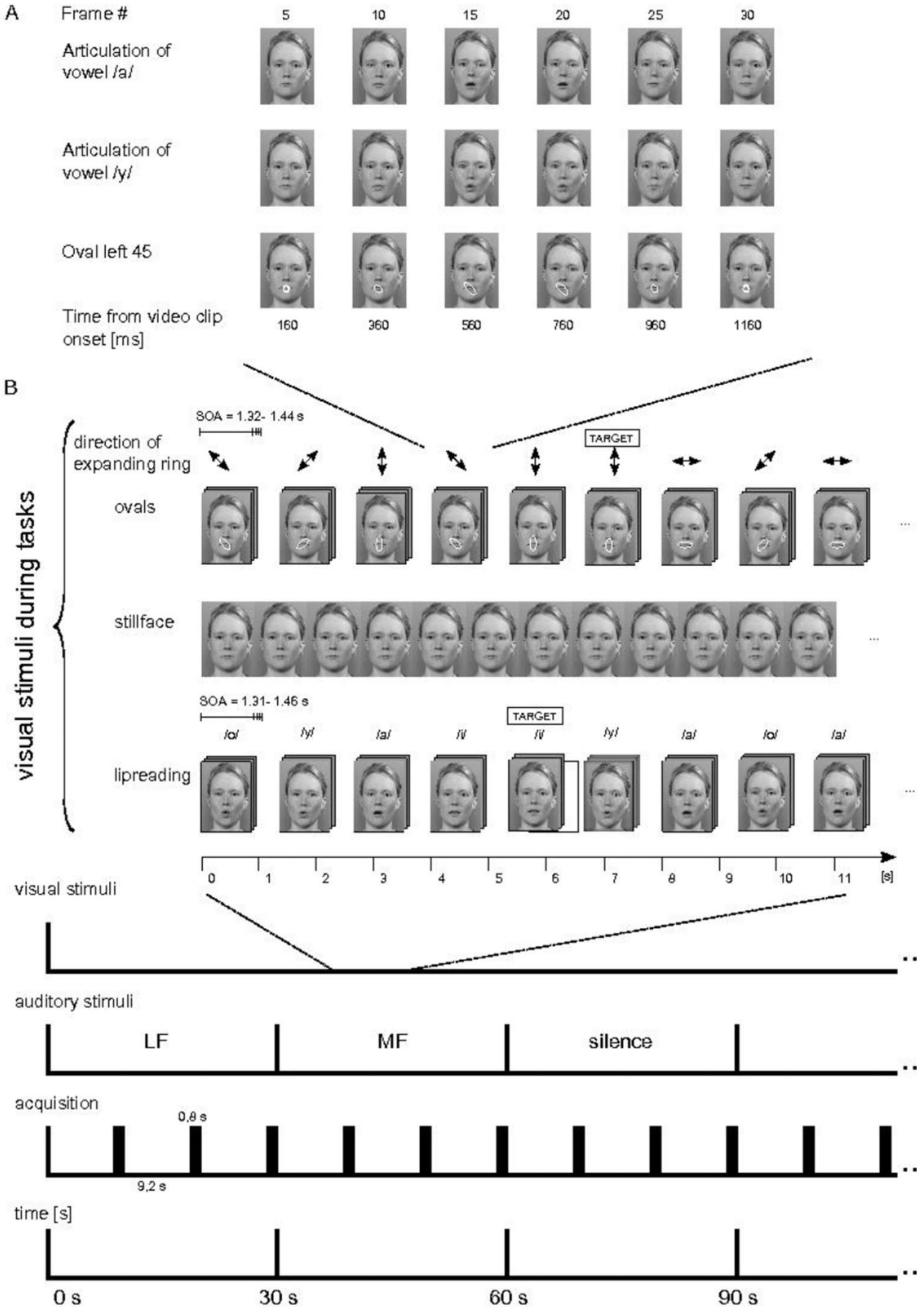

Figure 1.

Schematic illustration of stimuli, experimental paradigm, and functional magnetic resonance imaging (fMRI) acquisition. **A**) Still images from three example video clips showing Finnish vowel /a/ and /y/ articulations and an expanding circle with a final position of left 45° oblique oval are shown at constant intervals. The video clips of the visual conditions were edited to be comparable in timing and spatial frequency; also a circle transformation into oval resembled the mouth opening in the vowel articulation. **B**) The subjects were shown, in different conditions, an expanding circle into an oval of four alternative directions, still face, and articulation of Finnish vowels /o/, /y/, /a/, and /i/. The subjects were performing a one-back task where consecutive repetition of a stimulus constituted a target (*i.e.*, when a given circle expansion direction was repeated consecutively, or when a given vowel was presented in a row). During presentation of the expanding circle, still face, and vowel articulations, there was background stimulation either with low frequency (LF; center frequency at 250 Hz) sounds, mid-frequency (MF; center frequency at 2000 Hz) sounds, or silence. The functional echo planar imaging (EPI) acquisitions of 0.8 s were performed at intervals of 9.2 s using sparse sampling technique. Each stimulus block lasted for 30 s at a time.

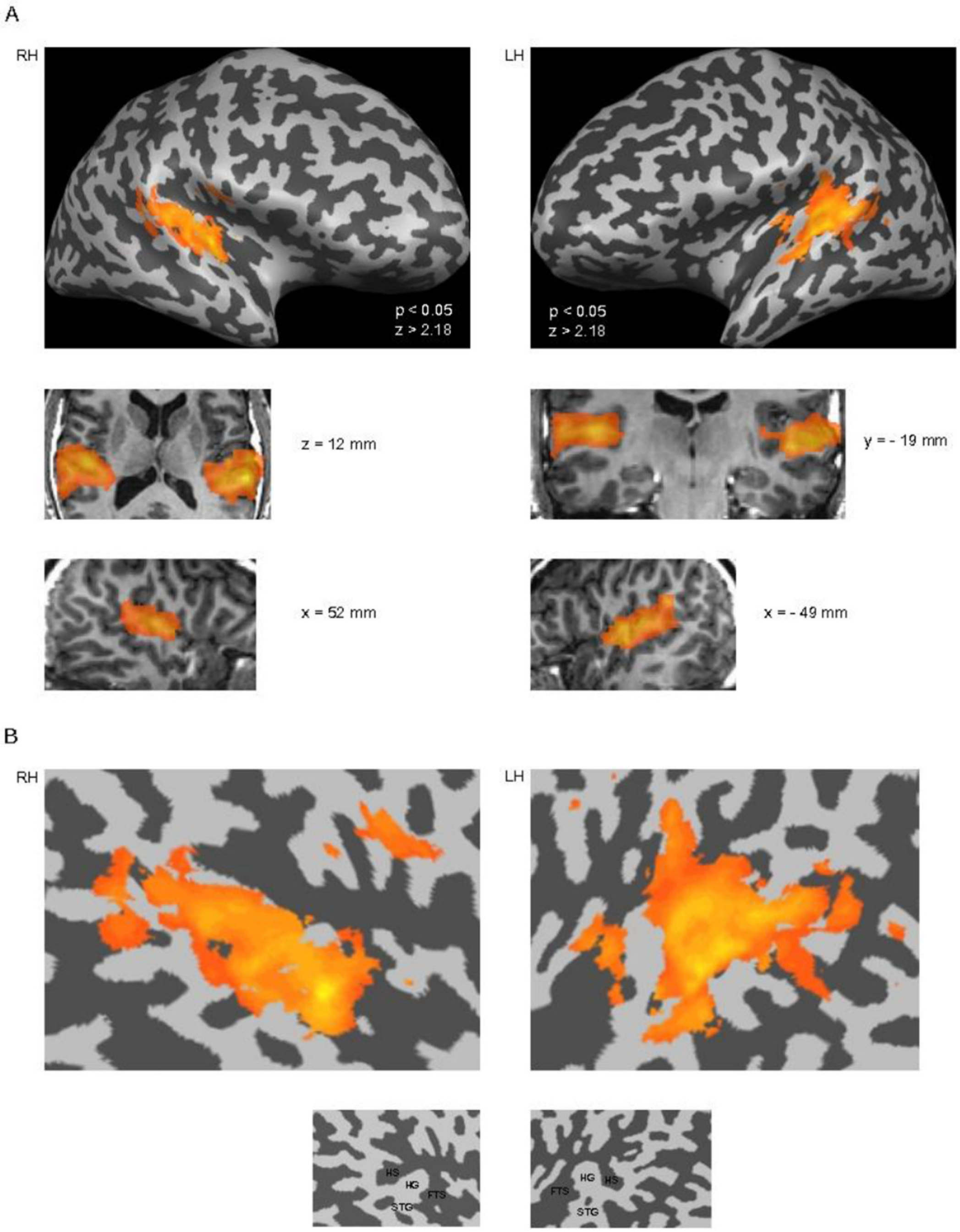

Figure 2.

Regions of interests (*i.e.*, masks) in the superior temporal lobes. The masks were obtained by contrasting the LF noise burst and MF noise burst stimulus conditions, as separate predictors, against silence in each condition at the significance level P < 0.05. The resulting activation maps were combined to form superior temporal lobe masks for each hemisphere, wherein the specific planned contrasts between the conditions were carried out. **A**) The masks are overlaid on the inflated brain of a single subject on each side of the image (gyri are colored light gray and sulci dark grey). The right hemisphere (RH) representation is on the left of the figure and the left hemisphere (LH) representation is on the right side of the figure. The middle two rows display the masks on selected transverse, coronal, and sagittal anatomical slices in standard Talairach space. **B**) The masks are presented on unfolded temporal-cortex patches of a single subject. As can seen, the masks encompassed extensive areas on the superior aspects of the temporal lobes bilaterally. (HS = Heschl's sulcus, HG = Heschl's gyrus, FTS = first transverse sulcus (a.k.a. lateral sulcus), STG = Superior temporal gyrus).

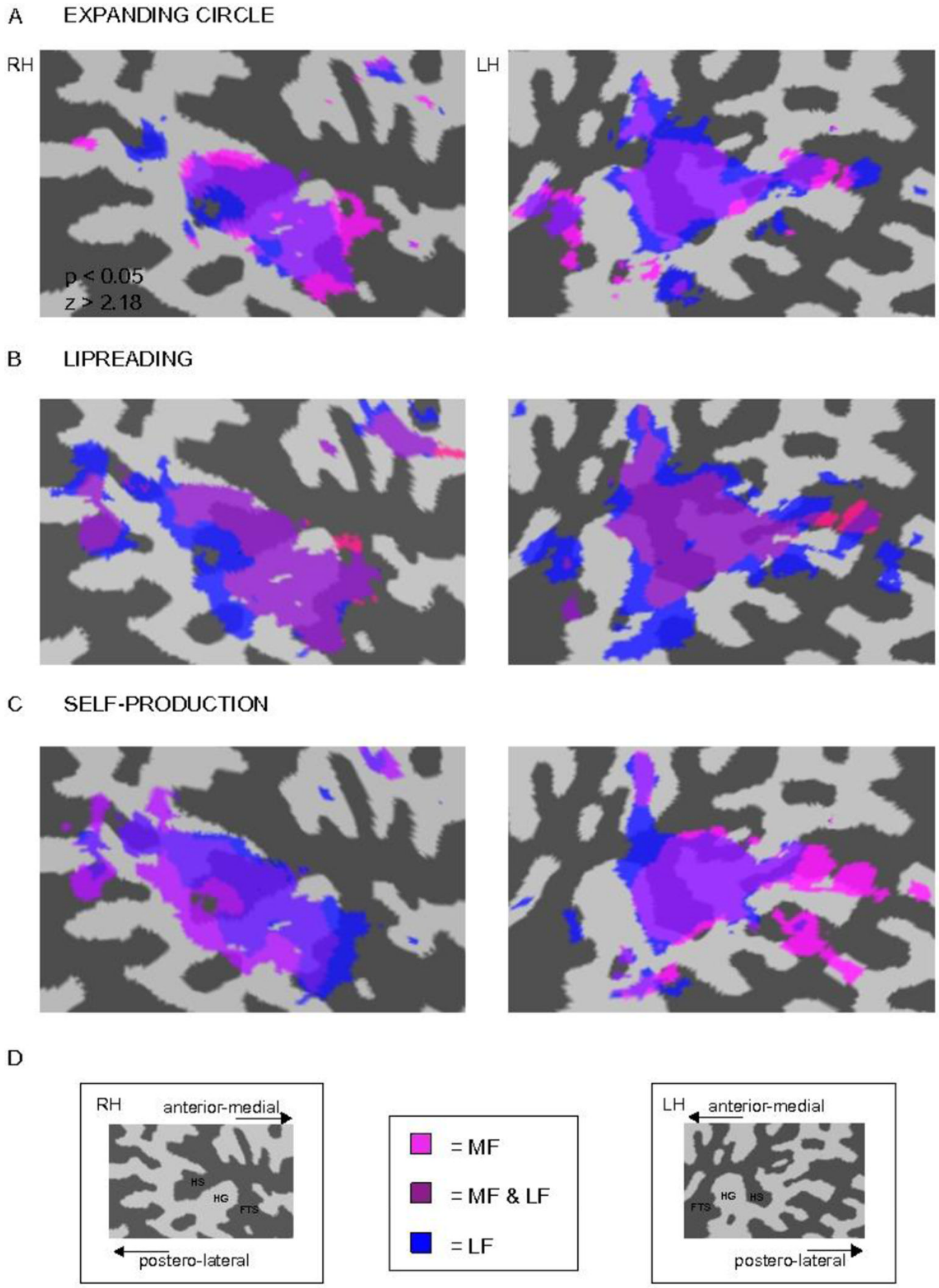

Figure 3.

Loci of hemodynamic activity in each of the conditions (*vs.* silence) on a single-subject right and left hemisphere temporal-cortex patches. **A**) Activity during perception of the expanding circle, **B**) during lipreading, and **C**) during self-production are shown. The RH representation is on the left of the figure and LH representation is on the right side of the figure. LF noise burst and MF noise burst were contrasted as separate predictors against silence in each condition at the significance level P < 0.05. **D**) The labeled cortical temporal-cortex patches of each hemisphere. The activations of each noise burst in different conditions are indicated with color coding in the middle (MF = pink, LF = blue, and areas activated by both LF and MF = magenta).

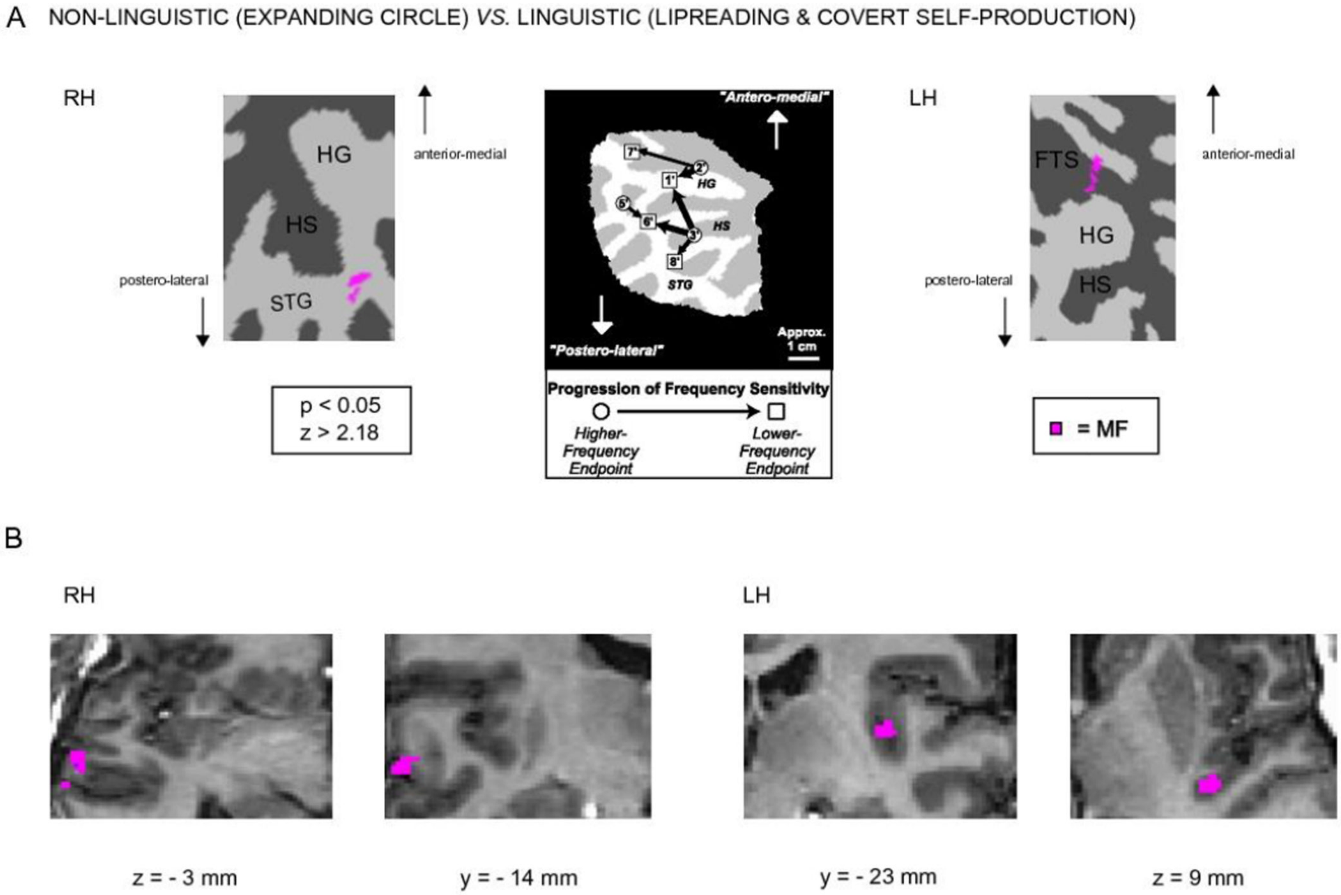

Figure 4.

Activations in the non-linguistic (expanding circle) *vs.* the linguistic conditions (lipreading and self-production) on the left and right hemisphere temporal-lobe cortex patches. **A**) In the right hemisphere, significantly stronger hemodynamic activity was observed in the STG lateral to HS in the non-linguistic condition, and in the left hemisphere there was stronger activity in the FTS anterior to HG in the non-linguistic condition. In the middle is shown tonotopic progressions adapted with permission from [40] for visual comparison; the left hemisphere and right hemisphere activations observed in the present study seem to correspond to the high-frequency end-points #2' and #5', respectively, of [40]. **B**) The activations are shown on selected transaxial and coronal anatomical slices in standard Talairach space.