Audiovisual Integration During Speech Comprehension: An fMRI Study Comparing ROI-Based and Whole Brain Analyses

Gregor R. Szycik,[1,2] Henk Jansma,[1] and Thomas F. Münte[1,3]*

[1] *Department of Neuropsychology, University of Magdeburg, Magdeburg, Germany*
[2] *Department of Psychiatry, Medical School Hannover, Hannover, Germany*
[3] *Center for Behavioral Brain Sciences, Magdeburg, Germany*

Abstract: Visual information (lip movements) significantly contributes to speech comprehension raising the question for the neural implementation of audiovisual (AV) integration during speech processing. To replicate and extend earlier neuroimaging findings, we compared two different analysis approaches in a slow event-related fMRI study of healthy native speakers of German who were exposed to AV speech stimuli (disyllabic nouns) with audio and visual signals being either congruent or incongruent. First, data was subjected to whole brain general linear model analysis after transformation of all individual data sets into standard space. Second, a region of interest (ROI) approach based on individual anatomy was used with ROI defined in areas identified previously as being important for AV processing. Standard space analysis revealed a widespread cortical network including the posterior part of the left superior temporal sulcus, Broca's region and its right hemispheric counterpart showing increased activity for incongruent stimuli. The ROI approach allowed to identify differences in activity between Brodmann areas 44 and 45, within Broca's area for incongruent stimulation, and also allowed to study activity of subdivisions of superior temporal regions. The complementary strengths and weaknesses of the two analysis approaches are discussed. *Hum Brain Mapp 30:1990–1999, 2009.* © 2008 Wiley-Liss, Inc.

Key words: speech; audiovisual; fMRI; multimodal processing

INTRODUCTION

One of the major tasks of an organism is the mapping of information from different modalities into a coherent representation of the environment. A special case of such multisensory integration processes is presented by stimuli that are inherently multimodal in nature, such as the perception of speech which is not only characterized by the auditory information associated with an utterance but also by the characteristic lip movements associated with each phoneme. Although this information may seem redundant in most cases, incomplete information from one modality can be compensated by the other. Thus, speech comprehension in noisy environments is considerably improved when the comprehender can look at the speaker's articulatory movements [Ross et al., 2007; Sumby and Pollack, 1954].

The presentation of audiovisual (AV) incongruent (auditory stream does not match the articulatory movements) speech stimuli may also lead to novel percepts that neither match the auditory nor the visual information as

Contract grant sponsor: BMBF; Contract grant number: 01GO0202; Contract grant sponsor: Deutsche Forschungsgemeinschaft SFB TR 31.

*Correspondence to: Thomas F. Münte, MD, Department of Neuropsychology, University of Magdeburg, Universitätsplatz 2, Gebäude 24, 39106 Magdeburg, Germany.
E-mail: thomas.muente@medizin.uni-magdeburg.de

Received for publication 22 January 2008; Revised 15 June 2008; Accepted 24 June 2008

DOI: 10.1002/hbm.20640

Published online 18 August 2008 in Wiley InterScience (www.interscience.wiley.com).

evidenced by the McGurk effect [McGurk and MacDonald, 1976]. This effect illustrates that the visual modality information in speech processing is more than a mere source of redundant information even in normal nonhearing impaired listeners. Behavioral studies comparing responses to semantically and/or spatially congruent or incongruent multisensory inputs to their unimodal counterparts have demonstrated a facilitatory influence of congruent stimuli on reaction times, whereas incongruent inputs led to increased response times compared with unimodal stimuli [Frens and van Opstal, 1995; Hershenson, 1962; Morrell, 1968; Sekuler et al., 1997; Stein et al., 1989]. In this investigation, we employ AV speech stimuli (bisyllabic words) for which the auditory and visual information either match (same word in both channels) or mismatch (visual word different from auditory word). Although one might object that the use of AV incongruent speech stimuli of this kind constitutes an unnatural situation, the rationale behind this manipulation is twofold. First, the use of violation stimuli is a standard practice in psycholinguistics, as the brain's reaction to these violations is presumed to reflect the processing of the information type that is violated. Second, and more importantly, a human listener might be confronted with AV incongruent speech quite often. Consider, for example, the typical cocktail-party situation [Cherry, 1953] which might involve not only several auditory speech streams but also several faces with lip movements. If lip movements of a particular face do not match the current auditory input, this might be an important additional clue to reject this particular auditory message [Devergie et al., 2008]. We therefore see the use of AV congruent and incongruent speech stimuli as a means towards defining the role of visual information in disambiguating complex auditory scenes by providing information for the enhancement (in the case of congruent AV speech) or rejection (in the case of incongruent AV speech) of a speech message. To summarize, the comprehension of language is obviously the result of the integration of auditory and visual modality information that is used and combined in a flexible manner.

Although the behavioral data are clear-cut, the question arises as to where and how the brain is accomplishing AV integration during speech comprehension. Depending on the experimental settings and methods used, different cortical areas appear to be involved in such AV speech processing [for reviews see: Calvert, 2001; Calvert and Thesen, 2004]. There is good evidence for participation of heteromodal cortex centered upon the sulcus temporalis superior (STS) [Beauchamp et al., 2004a,b; Reale et al., 2007; Szycik et al., 2007]. The caudal part of STS revealed different activation for AV speech compared with their unimodal components presented separately. This change in activity was shown for syllables [Sekiyama et al., 2003], for monosyllabic words [Wright et al., 2003], and for speech sequences [Calvert et al., 2000]. In addition, this region has been shown to be responsive to graphemic (letter) stimuli [Raij et al., 2000; van Atteveldt et al., 2004]. The visual component of AV speech stimuli exerts a modulatory influence on the auditory areas located in the dorsal surface of the temporal lobe [Callan et al., 2001; Calvert et al., 1999; Möttönen et al., 2004]. Indeed, the primary auditory cortex (PAC) appears to be involved in visual speech perception [Calvert et al., 1997; Pekkola et al., 2005]. But also speech relevant areas show activity differences as a function of AV stimulation. For example, the presentation of AV incongruent vowels (e.g. auditory/a/and visual/y/) compared with AV congruent vowels (both modalities/a/) is associated with greater activity in Broca's region [Ojanen et al., 2005]. To summarize, a range of brain regions including the STS, Broca's area, as well as primary and secondary auditory cortex have been suggested to be involved in the processing of AV speech.

A potential problem with most previous neuroimaging studies in this field is that they employed analysis strategies in standard space. Although this approach has been shown to be quite sensitive, it has also been pointed out that grand averages of brain transforms tend to blur and mislocalize activations in the cortex because of large interindividual anatomical variability. Such variability is a well-known problem in auditory areas [Penhune et al., 1996; Rademacher et al., 2001] but has also been pointed out for inferior frontal cortex, that is Broca's area [Amunts et al., 1999].

We therefore decided to re-examine the functional neuroanatomy of AV integration during speech comprehension by using and comparing two analysis approaches: In addition to a general linear model (GLM) analysis in standard Talairach space, [Talairach and Tornoux, 1988] we employed a region of interest (ROI) approach. For the latter we defined, based on individual anatomical landmarks, ROIs in areas that have been shown previously to play a potential role in AV speech perception. This approach obviously requires a hypothesis-driven definition of cortical regions, and therefore has to be restricted to a few areas. We were particularly interested to what extent we could find subdivisions in superior temporal sulcus and superior temporal gyrus as well as in Broca's area that are differentially sensitive to AV integration processes. Functional specialization within Broca's has been suggested by previous neuroimaging studies, [Bookheimer, 2002; Cannestra et al., 2000; Gelfand and Bookheimer, 2003; Hagoort, 2005] and was therefore of interest also with regard to AV processing of speech [Ojanen et al., 2005; Pekkola et al., 2006]. By using both methods in parallel, it is possible to benefit from the anatomical accuracy of the ROI approach without losing the sensitivity of the whole brain analysis in standard space.

METHODS

Participants

Twelve healthy German native speakers (5 females, mean age 24.6 ± 2.1, range 21–29) participated after having given informed written consent. All procedures have been approved by the ethics committee of the University of Magdeburg. Data from one subject was excluded because of self-reported ambiguous handedness. All remaining participants were right-handed.

Stimuli and Design

Stimuli were derived from the German part of the CELEX-Database [Baayen et al., 1995], and comprised 70 disyllabic nouns with a Mannheim frequency (1,000,000) of at least one. The stimuli were spoken by a female German native speaker with linguistic experience, and recorded by means of a digital camera and a microphone. The video was cut into 2 s segments (400^2 pixel resolution), showing the frontal view of the whole speakers face as she spoke a word. The accompanying audio stream was in mono-mode (used software: Adobe Premiere 6.5 for video processing and Adobe Audition 1.0 for audio processing). The stimuli were randomly divided into two sets of 35 items each. The first set contained video sequences with congruent audio–visual (AV) information (lip movements fitting to the spoken word; AV-congruent condition). The second set comprised video sequences with incongruent information in the audio and video channels (lip movements did not fit to the spoken word; AV-incongruent condition; e.g., video: Insel/island, audio: Hotel/hotel). The incongruent stimuli were created by randomly mixing the video and audio stream of the movies from the second set.

A slow event-related design was used for the stimulus presentation. Each stimulation period (2 s) was followed by a 16 s resting period with a fixation cross at the position of the speaker's mouth. To keep participants attending to the stimuli, they were required to identify words belonging to a specific semantic target category (i.e. animals, total number of occurrences four) by pressing one of two buttons with the left/right index finger depending on whether a target was present/not present. The participants had thus to respond for each stimulus. The responses to the targets were discarded from further analysis. As the responses to the critical nontargets required always a right index finger movement, activity related to the motor response was equated for AV congruent and incongruent stimuli and, thus is cancelled out in the contrast of interest.

Presentation software (Neurobehavioral Systems) was used to deliver stimuli. Stimuli were presented via fMRI compatible electrodynamic headphones integrated into earmuffs for reduction of residual background scanner noise [Baumgart et al., 1998]. The sound level of stimuli was individually adjusted to good audibility. Visual stimuli were projected via a mirror system by LCD projector onto a diffusing screen inside the magnet bore.

Image Acquisition

Magnetic-resonance images were acquired on a 3T TRIO Siemens scanner (Erlangen, Germany) equipped with a standard head coil. A total of 650 T_2^*-weighted volumes of the whole brain (TR 2000 ms, TE 30 ms, flip angle 80°, FOV 224 mm, matrix 64^2, 30 slices, slice thickness 3.5 mm, and interslice gap 0.35 mm) near to standard bicommisural (ACPC) orientation were collected. After the functional measurement, T_1-weighted images (TR 1550 ms, TE7.3 ms, flip angle 70°, FOV 224 mm, and matrix 256^2) with slice orientation identical to the functional measurement were acquired to serve as a structural overlay. Additionally, a 3D high resolution T_1-weighted volume for cortex surface reconstruction (FLASH, TR 15 ms, TE 4.9 ms, flip angle 25°, matrix 192×256^2, and 1 mm isovoxel) was recorded. The subject's head was fixed during the entire measurement to avoid head movements.

fMRI Data Analysis

First the subject's head motion was detected by using Brain Voyager QX software (rejection criterion: head translation of more than 1 mm or rotation of more than 1°) Three datasets passed the exclusion criterion. The remaining eight datasets were motion- and slice scan time corrected before further analysis. Additional linear trends and nonlinear drifts were removed by temporal filtering. Finally, after the coregistration with the structural data, a spatial transformation into the standard Talairach space [Talairach and Tornoux, 1988] was applied. The data was statistically analyzed both in standard space and by mean of a ROI approach.

Analysis of activation in individual ROIs

Seven auditory ROIs were defined on the basis of macro-anatomical landmarks in each hemisphere (e.g., the individual ROIs in the left hemisphere see Fig. 1). On the dorsal surface of the temporal lobe, the PAC and secondary belt regions T2 (covering the concave part of Heschl's sulcus including the posterior wall of Heschl's gyrus) and T3 (approximately covering the whole planum temporale) were defined. T2 and T3 areas have been described in more detail by Brechmann et al. [2002]. PAC was defined as the anterior convex part of the medial third of the Heschl's gyrus (first transverse temporal gyrus), corresponding to the microanatomically defined area KAm of Galaburda and Sanides [1980] and Te1.0 of Morosan et al. [2001].

Two ROIs were defined on the lateral surface of the temporal lobe: The ROI STS (superior temporal sulcus) was delineated on the concave surface of this macroanatomically prominent structure in accordance with the definition by Ochiai et al. [2004]. Dorsal from the STS we defined the convexity of the superior temporal gyrus as ROI STG.

Finally, we defined two ROIs on the lateral surface of the frontal lobe corresponding to the Brodmann's areas 44 and 45 [Brodmann, 1909] covering Broca's region in the left hemisphere. The macroanatomical landmarks used for the identification of the frontal ROIs were derived from Amunts et al. [1999] and Tomaiuolo et al. [1999]. Thus, for

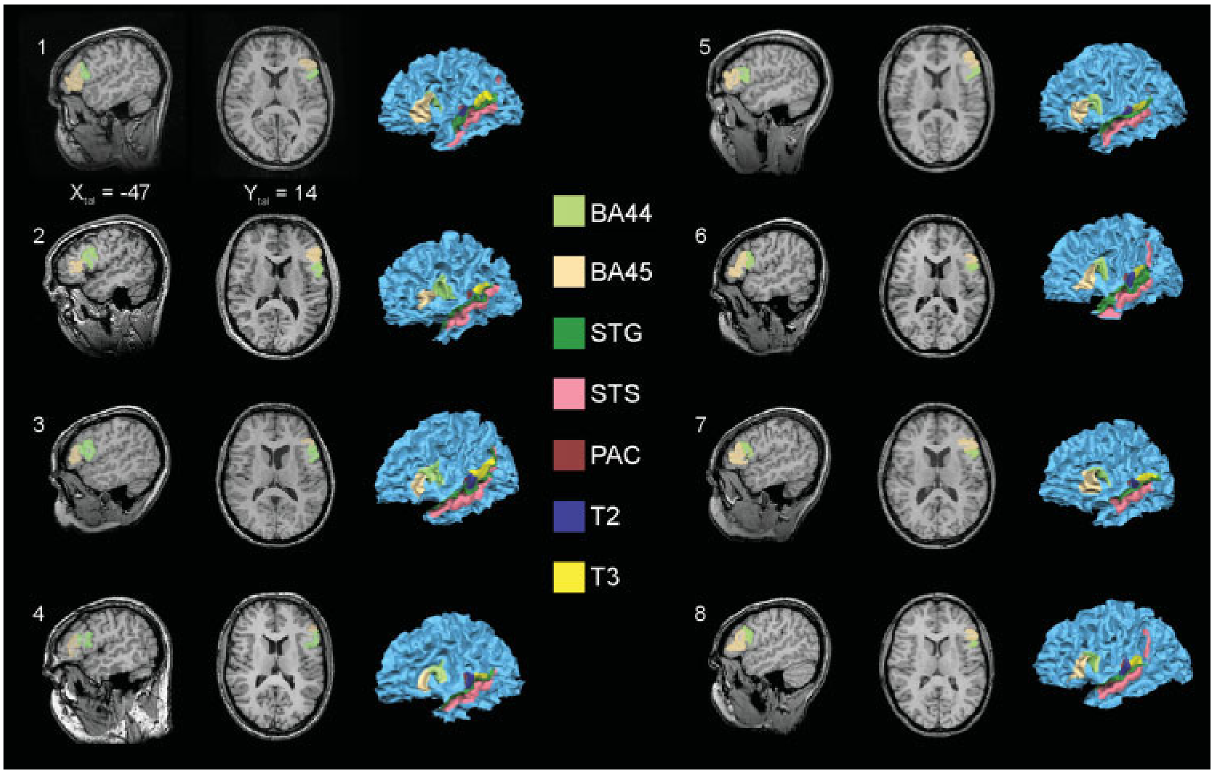

Figure 1.

Illustration of left hemispheric individual ROIs. Depicted are one sagittal and one horizontal slice of each participant's brain and a three-dimensional reconstruction of the white/gray matter border of the left hemisphere. For each individual participant Brodmann areas 44 and 45 (BA44; BA45) of Broca's region are shown in the slices and projected on the 3-D structural data. In addition all other left hemispheric ROIs are shown on the surface representation. We defined the individual ROIs using the reconstructed image of the hemisphere by tracing the ROI borders on the individual surface with continuous checking the position on the individual structural 3D volume. This kind of ROI definition is more reliable and comfortable than the use of only the 3D volume data, because the simultaneous view on the reconstructed brain allows a better identification of macroanatomical landmarks like specific gyri or sulci. Each point of the surface is coregistered with a specific voxel of the 3-D volume. X_{tal} Talairach coordinate X, Y_{tal} Talairach coordinate Y, color-coded are the specific ROIs. See text for more details of ROI definition.

BA44 the dorsal border was the fundus of the inferior frontal sulcus, the caudal border was defined by the fundus of the sulcus praecentralis. The border between BA44 and BA 45 depended on the presence of the ramus ascendens of the sulcus lateralis or the sulcus diagonalis or, in case that both sulci were present, a virtual line central between them. The ventrorostral border of BA45 was located in the fundus of the ramus horizontalis of the sulcus lateralis. Both ROIs did not extend into the deepness of the Sylvian fissure.

For each individual ROI, significant voxels were detected by means of a GLM. For the statistical analysis, we used functional data acquired during the presentation of all events independent of the participant's response. In the GLM, we defined a hemodynamic response function for each experimental condition by convolving the box-car function with the model of Boynton et al. [1996] using δ = 2.5, τ = 1.25, and n = 3. The false discovery rate threshold of q (FDR) < 0.05 [Genovese et al., 2002] was chosen for identification of the activated voxels. Subsequently, the time course of the BOLD response averaged over all significant voxels from each subject-specific ROI was extracted and z-transformed. The calculated time series of corresponding ROIs were averaged over subjects resulting in a mean time-course depicting the BOLD answer for each ROI and experimental condition. The values of the third and fourth time point after the stimulation (corresponding to the BOLD maximum, 6 and 8 s after stimulus onset) were used to test for differences between the experimental conditions by means of t-tests.

Because of the large size of the STG and STS ROIs and the fact that these regions definitely harbor different functional areas, we decided to subdivide them into smaller parts. Both ROIs were therefore traced on 13 slices (every 5 mm one slice) in frontal orientation ranging from 0 to 60 mm posterior from the anterior commissure. This tech-

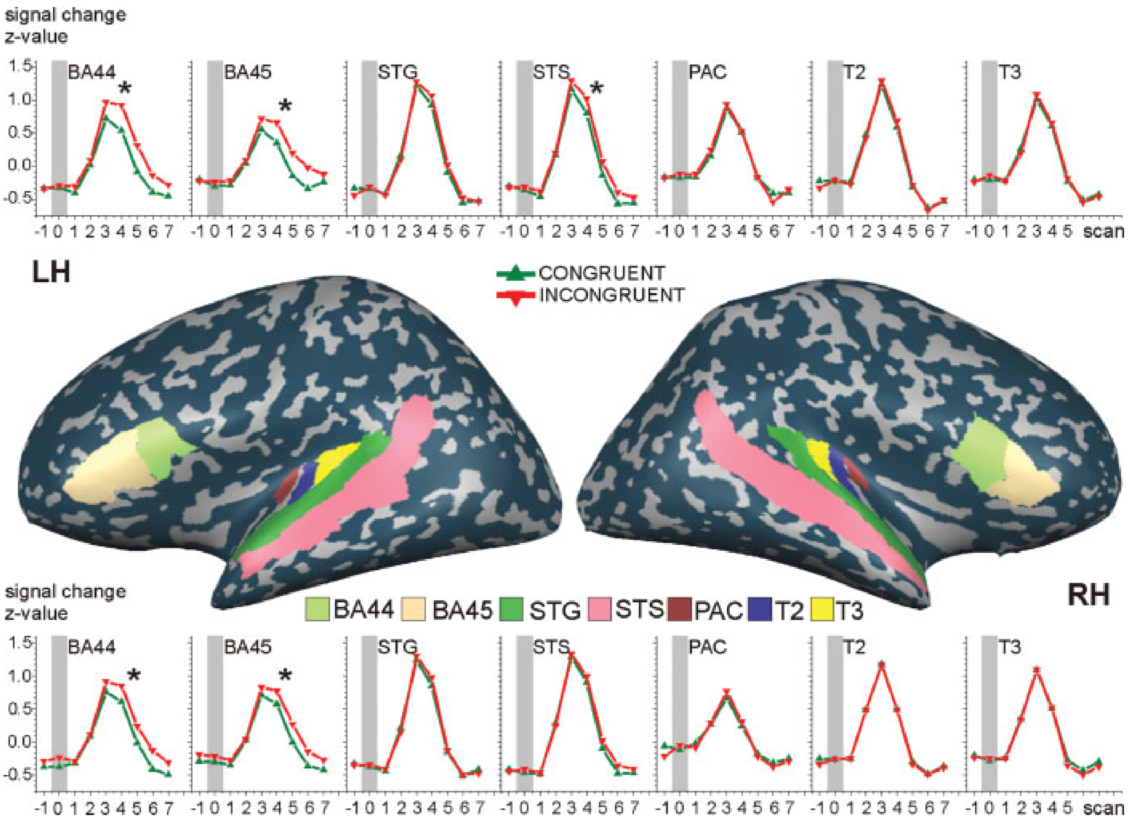

Figure 2.

BOLD-time courses for the left (LH) and right hemisphere (RH). There are significant stronger signal for audiovisual incongruent condition (red line) in comparison to the congruent one (green line) in both hemispheres in BA44 and BA45. In addition left hemispheric STS showed the same difference. For the purpose of visualization, left and right partially inflated lateral views on the hemispheres of one volunteer are depicted. ROIs are color coded. BA44, Brodmann area 44; BA45, Brodmann area 45; STG, superior temporal gyrus; STS, superior temporal sulcus; PAC, primary auditory cortex; T2 and T3 auditory areas on the dorsal surface of the temporal lobe. Gray bars indicate stimulus presentation. Asterisks show significant differences.

nique is similar to the one described by Wright et al. [2003]. For statistical analysis only the data from those slices was used that contained significantly activated voxels in each subject. The group analysis strategy was the same as described for not subdivided ROIs.

Analysis of activation in Talairach standard space

To identify possible regions of activity outside the predefined ROIs, group data were analyzed by multi-subject fixed-effects GLM in standard space. The predictors used were derived in the same way as described for the ROI analysis. To emphasize spatially coherent activation patterns, functional data was additionally spatially smoothed with a Gaussian kernel of 10 mm full width at half maximum. We calculated two statistical contrasts on significant voxels at q (FDR) < 0.05 (stimulation vs. baseline). The first contrast revealed voxels that showed stronger signal changes for the AV incongruent condition in comparison with the AV congruent condition, whereas the second comparison reflected the opposite contrast.

RESULTS

Analysis of Activation in Individual ROIs

The averaged time course of each ROI showed a typical BOLD answer with the maximum signal change occurring about 6 s after the stimulus onset (Fig. 2). Three ROIs in the left hemisphere showed significantly stronger signal changes in the incongruent, that is BA44 ($P < 0.001$, Cohen's effect size $D = 0.422$), BA45 ($P < 0.001$, $D = 0.289$) and STS ($P < 0.001$, $D = 0.267$). Two ROIs in the right hemisphere similarly showed stronger activation in the incongruent condition: BA44 ($P < 0.003$, $D = 0.263$) and BA45 ($P < 0.014$, $D = 0.212$). None of the selected ROIs showed a stronger activation to the congruent stimuli.

All slices of the *right* STG in the range from AC to 35 mm showed significantly activated voxels in all partici-

TABLE I. Summary of the results of the analysis in standard space

		Talairach coordinates					
Cluster	Volume (mm^3)	*X*	*Y*	*Z*	BA (distance, mm)	Involved BA	Volume (mm^3)
LH							
Cluster 1	689	−63	−32	−3	21 (3)	21	300
						22	9
Cluster 2	811	−53	−54	8	39 (5)	39	87
						22	27
						21	19
						37	9
Cluster 3	7352	−49	14	18	44 (3)	9	931
						44	793
						45	694
						6	328
						47	73
						13	44
						8	2
Cluster 4	1908	−31	18	−3	47 (1)	13	239
						47	215
Cluster 5	3803	−6	13	49	6 (3)	6	1199
						8	278
						32	44
						24	33
Cluster 6	641	−40	−59	45	40 (5)	7	171
						40	72
Cluster 7	664	−4	34	33	9 (3)	9	150
						6	115
						32	36
						8	4
RH							
Cluster 8	392	46	33	7	46 (7)	46	14
Cluster 9	2900	45	16	22	9 (3)	9	288
						44	85
						8	82
						46	59
						13	55
						45	31
						6	7

Given is the volume in mm^3 of each cluster, the Talairach coordinates and corresponding Brodmann area (BA) of its center of mass. In parentheses the distance (in mm) between the center point and the nearest anchor point that refers to a specific BA in Talairach. All BAs that share volume with a particular cluster are also given.

pants but no differences due to the experimental conditions. For the right STS, a similar pattern was observed but between AC and the slice at 45 mm. The left STG showed significantly activated voxels from the AC-line to slice 40 mm. In addition, there was significantly stronger activation ($P < 0.030$, $D = 0.182$) for the AV incongruent condition in the slice 10 mm. For the left STS significantly activated voxels were observed from slice 10 mm to slice 40 mm. The incongruent condition led to significantly stronger signal changes in the slices 25 mm ($P < 0.028$, $D = 0.169$), 35 mm ($P < 0.013$, $D = 0.196$), and 40 mm ($P < 0.003$, $D = 0.236$).

Analysis of Activation in Standard Space

The analysis of the GLM contrasts revealed several activity clusters, seven in the left and two in the right hemisphere (for details, see Table I), showing stronger signal changes for incongruent in comparison to congruent stimuli in both hemispheres (Fig 3). Left hemispheric activations involve the medial and superior temporal cortex (BA 21 and 22, Cluster 1), posterior temporal cortex (BA 39, 22, 21, 37, and Cluster 2), lateral prefrontal cortex (BA 9, 44, 45, 6, 47, 13, and 8, Cluster 3), insula and inferior prefrontal cortex (BA 13 and 47, Cluster 4), medial prefrontal cortex (BA 6, 8, 32, and 24, Cluster 5; BA 9, 6, 32, and 8, Cluster 7), and parietal cortex (BA 7 and 40, Cluster 6).

In the right hemisphere, activation was seen in dorsolateral prefrontal cortex (BA 46, Cluster 8) and in lateral prefrontal cortex (BA 9, 44, 8, 46, 13, 45, and 6, Cluster 9).

As for the ROI analysis, no significant activations were found for the contrast congruent versus incongruent stimuli

DISCUSSION

The present investigation aimed at comparing the standard whole-brain GLM analysis in standard space with a

ROI-based approach resting on a selected number of predefined regions of interest that were determined individually on each participants anatomy.

ROI-Based Approach

All of the predefined ROIs contained significant active voxels for the contrast stimulation versus rest. Early auditory areas PAC, T2, and T3 [corresponding to core, belt, and partially parabelt regions of the primate [Pandya, 1995; Semple and Scott, 2003] showed a ramp-like increase and a fast decrease of the BOLD-curve, which is clearly different from the areas BA44 or BA45, which showed a plateau. There was no differential activity for congruent and incongruent AV stimuli for the early auditory areas, suggesting that these are not dealing with the specific aspect AV speech processing addressed in this study. This is in contrast to earlier electrophysiological studies that demonstrated [Klucharev et al., 2003; Möttönen et al., 2004] involvement of auditory areas. However, rather than presenting congruent and incongruent AV stimuli, these studies employed auditory only and visual only syllable stimuli and compared the associated neuromagnetic responses with phonetically congruent AV syllables.

Our study revealed left hemispheric specialization of the caudal part from STS for the AV processing of speech stimuli. The caudal part of STS has been identified as a multisensory integrative site involved in processing auditory as well as visual stimuli [Beauchamp et al., 2004a,b; Callan et al., 2003; Wright et al., 2003]. In particular, this region seems to be involved in the processing of AV speech [Beauchamp, 2005; Bernstein et al., 2008; Szycik et al., 2007]. Calvert et al. [2000] found in the vicinity of the STS, a functional activity cluster with stronger responses for congruent AV speech stimuli. In contrast, we did not find stronger STS activations for matching than conflicting AV speech stimulation. Our results agree with recent studies [Ojanen et al., 2005; Pekkola et al., 2006], and raise the question for the origin of the contrary findings. The main difference between our study and that of Calvert et al. [2000] is the kind of stimulation. The stimuli of this study conflicted only with respect to the congruency of lip-movements, whereas that of Calvert et al. [2000] differed additionally in the temporal domain. Therefore following earlier suggestions, stronger activity for AV incongruent stimuli could be the result of the interaction between mirror neurons situated in Broca's region [Ojanen et al., 2005] and the neurons of the STS region via back projections [Hertrich et al., 2007; Nishitani et al., 2005; Skipper et al., 2007]. The involved neuronal populations may be different from those processing temporal synchronicity. Furthermore, Szycik et al. [2007] have shown different activity patterns in the STS for AV speech stimulation depending on whether or not noise was present in the auditory signal. During noiseless AV speech stimulation STS regions showed stronger activation for congruent stimuli, whereas under noisy conditions the same region showed stronger activation for incongruent AV speech. Apparently, in noisy environments the visual information affects speech processing to a greater extent. On the other hand, stronger activation for incongruent AV speech stimuli in this study may simply reflect increased effort of STS cell populations that show functional similarity to the Broca region mirror neurons. Indeed, an invasive study in rhesus macaques demonstrated neurons in the STS which form multisensory representations of seen actions [Barraclough et al., 2005].

In agreement with other studies [Ojanen et al., 2005; Pekkola et al., 2006] Broca's region and its right hemispheric homologue showed increased activation for incongruent AV speech. In this investigation, we found activations of both BA 44 and 45. Functional differences in the region of the left inferior frontal region in connection with language processing have been reported [Bookheimer, 2002]. In particular, Bookheimer, in her metaanalysis, pointed out that there is functionally defined gradient with semantic processes subserved by BA 47 and BA 45, syntax processing by BA 45 and BA 44, and phonological processing achieved by BA 44 extending into BA 6. In our brief stimuli, (disyllabic nouns) the "phonological" differences between the heard and the seen word of incongruent AV, words appear to be most relevant and have thus led to stronger activation of BA 44 as identified by the ROI analysis.

It has been shown that BA44 harbors mirror neurons analogue to those of the area F5 of monkey [Rizzolatti and Arbib, 1998] and therefore plays a crucial role in action understanding [Molnar-Szakacs et al., 2005]. Speech comprehension processes have been argued to have developed from those phylogenitically older processes of action observation [Arbib, 2005; Corballis, 2003; Rizzolatti and Arbib, 1998]. Such views have been supported by functional imaging studies [Aziz-Zadeh et al., 2006; Pulvermüller, 2005] as well as transcranial magnetic stimulation investigations [Meister et al., 2003; Tokimura et al., 1996] and are in line with the motor theory of speech [Liberman and Mattingly, 1985]. With regard to our results, we suggest that the mapping of the auditory and visual speech input onto the motor representation of articulatory movements underlies the activation of the BA44 subregion of Broca's area.

Standard Space Analysis

The whole brain analysis in standard space revealed activity in some of the functional clusters targeted by the ROI analysis but also in additional functional clusters outside the predefined areas. All of these showed stronger activation for incongruent stimuli. Our results are in agreement with other recent studies [Jones and Callan, 2003; Miller and D'Esposito, 2005; Ojanen et al., 2005]. As expected for speech stimuli, the majority of significant clusters were located in the left hemisphere with only two clusters found on the surface of the right hemispheric IFG (Fig. 3, Table I).

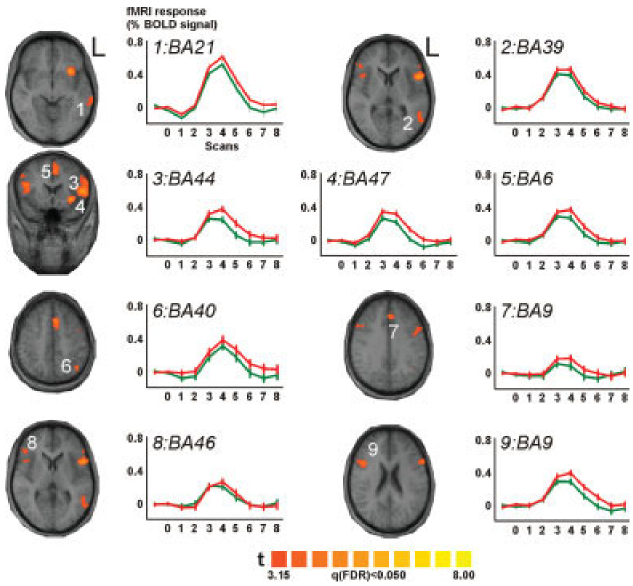

Figure 3.

Brain activations detected in the standard space group analysis with graphs depicting the BOLD-signal changes for the incongruent (red line) and congruent (green line) condition. The significant clusters are plotted on an average structural brain image constructed of the individual Talairach transformed 3D volumes. All show significantly stronger responses to the incongruent condition. The numbers correspond to the cluster numbers used in Table I and in the text. For each of the clusters the Brodmann area (BA) that corresponds to the center of mass of the particular cluster is given for each of the BOLD time-course graphs. L left.

The left hemispheric activity in Cluster 3 spans several functional areas. Beside Broca's region, parts of dorsolateral prefrontal cortex (DLPFC) fall into this cluster. The DLPFC plays an important role for planning and executing actions, for speech perception, and for perception of articulatory face movements [Skipper et al., 2005] which might explain its greater recruitment by AV incongruent stimuli.

Additional activation clusters were seen in the anterior insula region (Cluster 4), inferior parietal lobule (Cluster 6), and the mediofrontal cortex (Clusters 5 and 7). A key feature of incongruent stimuli is the mismatch between the articulatory lip movements and the variation of the auditory stream. The right anterior insula seems to be involved in processing of temporally dynamic information for nonspeech stimuli [Bushara et al., 2001]. Miller and D'Esposito [2005] have shown that the left anterior insula processes temporal correspondence of speech stimuli on a sensory level, and may be sensitive for the percept of AV fusion of the stimuli.

The involvement of the regions in the vicinity of intraparietal sulcus in processing of multimodal stimuli has been shown by several recent studies [Bushara et al., 2001; Calvert et al., 2001; Macaluso et al., 2004]. This region plays an important role for control of attention shifts to a sensory modality [Macaluso et al., 2000]. Such attentional shifts may be necessary for solving AV tasks with impoverished or faulty information in one modality as it is the case for AV incongruent speech [Jones and Callan, 2003; Miller and D'Esposito, 2005].

The mediofrontal cingulate cortex participates in the processing of tasks with variable degrees of difficulty [Corbetta et al., 1991; Paus, 2001; Paus et al., 1998], and supports important performance and conflict monitoring tasks [Carter et al., 1998, 1999, 2000; Ridderinkhof et al., 2004]. The activity of Cluster 5 and 7 may thus be a result of increased attentional and monitoring demands in the incongruent condition.

The right hemisphere is involved in the speech processing in multiple ways: it plays a crucial role in understanding humor, sarcasm, metaphors, or comprehension of prosody [Mitchell and Crow, 2005]. The right hemispheric homlogue of the Broca's region possesses speech relevant functions like involvement in imitation [Heiser et al., 2003] respectively inhibitory influence on certain imitatory responses [Nishitani et al., 2005]. The right hemispheric activation of IFG detected in our study may reflect involvement of this area in motor imitation processes during speech perception.

Comparison of the Two Approaches

Analysis in standard space clearly revealed areas involved in AV integration that had not been targeted by our hypothesis-driven ROI-approach, and thus might be considered to be the more sensitive analysis method. On the other hand standard space analysis leads to a decrease of spatial certainty and resolution because of the normalization and spatial smoothing procedures. In particular, this will make it difficult to distinguish distinct but adjacent functional areas. This short-coming of the standard space approach is evidenced by the finding of different effect sizes in left BA 44 and 45 suggesting a differential functional role of these two parts of Broca's area [see also: Bookheimer, 2002; Hagoort, 2005]. We suggest for future research, the increased combination of a hypothesis driven approach examining the activity in individually determined ROIs based on macroanatomical landmarks with a more exploratory approach in Talairach standard space particularly by increasing spatial accuracy using new standardization methods like the high-resolution cortical alignment that is based on the comparison of the curvature pattern of the cortical surface [Goebel et al., 2006]. Elsewhere, we have also shown the utility of a functional localizer approach in the analysis of AV integration in speech processing [Szycik et al., 2007].

ACKNOWLEDGMENTS

The authors thank Beate Wendt for patient and expert voicing of the stimuli.

REFERENCES

Amunts K, Schleicher A, Burgel U, Mohlberg H, Uylings HB, Zilles K (1999): Broca's region revisited: Cytoarchitecture and intersubject variability. J Comp Neurol 412:319–341.

Arbib MA (2005): From monkey-like action recognition to human language: An evolutionary framework for neurolinguistics. Behav Brain Sci 28:105–124.

Aziz-Zadeh L, Wilson SM, Rizzolatti G, Iacoboni M (2006): Congruent embodied representations for visually presented actions and linguistic phrases describing actions. Curr Biol 16:1818–1823.

Baayen RH, Piepenbrock R, Gulikers L (1995): The CELEX Lexical Database [CD-ROM]. Philadelphia: University of Pennsylvania, Linguistic Data Consortium.

Barraclough NE, Xiao D, Baker CI, Oram MW, Perrett DI (2005): Integration of visual and auditory information by superior temporal sulcus neurons responsive to the sight of actions. J Cogn Neurosci 17:377–391.

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

Beauchamp MS (2005): See me, hear me, touch me: Multisensory integration in lateral occipital-temporal cortex. Curr Opin Neurobiol 15:145–153.

Beauchamp MS, Argall BD, Bodurka J, Duyn JH, Martin A (2004a): Unraveling multisensory integration: Patchy organization within human STS multisensory cortex. Nat Neurosci 7:1190–1192.

Beauchamp MS, Lee KE, Argall BD, Martin A (2004b): Integration of auditory and visual information about objects in superior temporal sulcus. Neuron 41:809–823.

Bernstein LE, Auer ET Jr, Wagner M, Ponton CW (2008): Spatiotemporal dynamics of audiovisual speech processing. Neuroimage 39:423–435.

Bookheimer S (2002): Functional MRI of language: New approaches to understanding the cortical organization of semantic processing. Annu Rev Neurosci 25:151–188.

Boynton GM, Engel SA, Glover GH, Heeger DJ (1996): Linear systems analysis of functional magnetic resonance imaging in human V1. J Neurosci 16:4207–4221.

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

Brodmann K (1909): Vergleichende Lokalisationslehre der Großhirnrinde. Leipzig: Johann Ambrosius Barth.

Bushara KO, Grafman J, Hallett M (2001): Neural correlates of auditory-visual stimulus onset asynchrony detection. J Neurosci 21:300–304.

Callan DE, Callan AM, Kroos C, Vatikiotis-Bateson E (2001): Multimodal contribution to speech perception revealed by independent component analysis: A single-sweep EEG case study. Brain Res Cogn Brain Res 10:349–353.

Callan DE, Jones JA, Munhall K, Callan AM, Kroos C, Vatikiotis-Bateson E (2003): Neural processes underlying perceptual enhancement by visual speech gestures. Neuroreport 14:2213–2218.

Calvert GA (2001): Crossmodal processing in the human brain: Insights from functional neuroimaging studies. Cereb Cortex 11:1110–1123.

Calvert GA, Thesen T (2004): Multisensory integration: Methodological approaches and emerging principles in the human brain. J Physiol Paris 98:191–205.

Calvert GA, Bullmore ET, Brammer MJ, Campbell R, Williams SC, McGuire PK, Woodruff PW, Iversen SD, David AS (1997): Activation of auditory cortex during silent lipreading. Science 276:593–596.

Calvert GA, Brammer MJ, Bullmore ET, Campbell R, Iversen SD, David AS (1999): Response amplification in sensory-specific cortices during crossmodal binding. Neuroreport 10:2619–2623.

Calvert GA, Campbell R, Brammer MJ (2000): Evidence from functional magnetic resonance imaging of crossmodal binding in the human heteromodal cortex. Curr Biol 10:649–657.

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

Cannestra AF, Bookheimer SY, Pouratian N, O'Farrell A, Sicotte N, Martin NA, Becker D, Rubino G, Toga AW (2000): Temporal and topographical characterization of language cortices using intraoperative optical intrinsic signals. Neuroimage 12:41–54.

Carter CS, Botvinick MM, Cohen JD (1999): The contribution of the anterior cingulate cortex to executive processes in cognition. Rev Neurosci 10:49–57.

Carter CS, Braver TS, Barch DM, Botvinick MM, Noll D, Cohen JD (1998): Anterior cingulate cortex, error detection, and the online monitoring of performance. Science 280:747–749.

Carter CS, Macdonald AM, Botvinick M, Ross LL, Stenger VA, Noll D, Cohen JD (2000): Parsing executive processes: strategic vs. evaluative functions of the anterior cingulate cortex. Proc Natl Acad Sci USA 97:1944–1948.

Cherry EC (1953): Some experiments on the recognition of speech, with one and two ears. J Acoust Soc Am 25:975–979.

Corballis MC (2003): From mouth to hand: Gesture, speech, and the evolution of right-handedness. Behav Brain Sci 26:199–208.

Corbetta M, Miezin FM, Dobmeyer S, Shulman GL, Petersen SE (1991): Selective and divided attention during visual discriminations of shape, color, and speed: functional anatomy by positron emission tomography. J Neurosci 11:2383–2402.

Devergie A, Grimault N, Berthommier F, Gaudrain E, Healy EW (2008): Effect of lip movement cues on auditory streaming of concurrent speech. J Acoust Soc Am 123:3302.

Frens MA, Van Opstal AJ (1995): Spatial and temporal factors determine audio-visual interactions in human saccadic eye movements. Percept Psychophys 57:802–816.

Galaburda A, Sanides F (1980): Cytoarchitectonic organization of the human auditory cortex. J Comp Neurol 190:597–610.

Gelfand JR, Bookheimer SY (2003): Dissociating neural mechanisms of temporal sequencing and processing phonemes. Neuron 38:831–842.

Genovese CR, Lazar NA, Nichols T (2002): Thresholding of statistical maps in functional neuroimaging using the false discovery rate. Neuroimage 15:870–878.

Goebel R, Esposito F, Formisano E (2006): 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 27:392–401.

Hagoort P (2005): On Broca, brain, and binding: a new framework. Trends Cogn Sci 9:416–423.

Heiser M, Iacoboni M, Maeda F, Marcus J, Mazziotta JC (2003): The essential role of Broca's area in imitation. Eur J Neurosci 17:1123–1128.

Hershenson M (1962): Reaction time as a measure of intersensory facilitation. J Exp Psychol 63:289–293.

Hertrich I, Mathiak K, Lutzenberger W, Menning H, Ackermann H (2007): Sequential audiovisual interactions during speech perception: A whole-head MEG study. Neuropsychologia 45: 1342–1354.

Jones JA, Callan DE (2003): Brain activity during audiovisual speech perception: An fMRI study of the McGurk effect. Neuroreport 14:1129–1133.

Klucharev V, Mottonen R, Sams M (2003): Electrophysiological indicators of phonetic and non-phonetic multisensory interactions during audiovisual speech perception. Brain Res Cogn Brain Res 18:65–75.

Liberman AM, Mattingly IG (1985): The motor theory of speech perception revised. Cognition 21:1–36.

Macaluso E, Frith C, Driver J (2000): Selective spatial attention in vision and touch: Unimodal and multimodal mechanisms revealed by PET. J Neurophysiol 83:3062–3075.

Macaluso E, George N, Dolan R, Spence C, Driver J (2004): Spatial and temporal factors during processing of audiovisual speech: A PET study. Neuroimage 21:725–32.

McGurk H, MacDonald J (1976): Hearing lips and seeing voices. Nature 264:746–748.

Meister IG, Boroojerdi B, Foltys H, Sparing R, Huber W, Topper R (2003): Motor cortex hand area and speech: Implications for the development of language. Neuropsychologia 41:401–406.

Miller LM, D'Esposito M (2005): Perceptual fusion and stimulus coincidence in the cross-modal integration of speech. J Neurosci 25:5884–5893.

Mitchell RL, Crow TJ (2005): Right hemisphere language functions and schizophrenia: The forgotten hemisphere? Brain 128:963–978.

Molnar-Szakacs I, Iacoboni M, Koski L, Mazziotta JC (2005): Functional segregation within pars opercularis of the inferior frontal gyrus: Evidence from fMRI studies of imitation and action observation. Cereb Cortex 15:986–994.

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.

Morrell LK (1968): Temporal characteristics of sensory interaction in choice reaction times. J Exp Psychol 77:14–18.

Möttönen R, Schurmann M, Sams M (2004): Time course of multisensory interactions during audiovisual speech perception in humans: A magnetoencephalographic study. Neurosci Lett 363:112–115.

Nishitani N, Schurmann M, Amunts K, Hari R (2005): Broca's region: From action to language. Physiology (Bethesda) 20:60–69.

Ochiai T, Grimault S, Scavarda D, Roch G, Hori T, Riviere D, Mangin JF, Regis J (2004): Sulcal pattern and morphology of the superior temporal sulcus. Neuroimage 22:706–719.

Ojanen V, Mottonen R, Pekkola J, Jaaskelainen IP, Joensuu R, Autti T, Sams M (2005): Processing of audiovisual speech in Broca's area. Neuroimage 25:333–338.

Pandya DN (1995): Anatomy of the auditory cortex. Rev Neurol (Paris) 151:486–494.

Paus T (2001): Primate anterior cingulate cortex: Where motor control, drive and cognition interface. Nat Rev Neurosci 2:417–424.

Paus T, Koski L, Caramanos Z, Westbury C (1998): Regional differences in the effects of task difficulty and motor output on blood flow response in the human anterior cingulate cortex: A review of 107 PET activation studies. Neuroreport 9:R37–R47.

Pekkola J, Laasonen M, Ojanen V, Autti T, Jaaskelainen IP, Kujala T, Sams M (2006): Perception of matching and conflicting audiovisual speech in dyslexic and fluent readers: An fMRI study at 3 T. Neuroimage 29:797–807.

Pekkola J, Ojanen V, Autti T, Jaaskelainen IP, Mottonen R, Tarkiainen A, Sams M (2005): Primary auditory cortex activation by visual speech: An fMRI study at 3 T. Neuroreport 16:125–128.

Penhune VB, Zatorre RJ, MacDonald JD, Evans AC (1996): Interhemispheric anatomical differences in human primary auditory cortex: Probabilistic mapping and volume measurement from magnetic resonance scans. Cereb Cortex 6:661–672.

Pulvermüller F (2005): Brain mechanisms linking language and action. Nat Rev Neurosci 6:576–582.

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.

Raij T, Uutela K, Hari R (2000): Audiovisual integration of letters in the human brain. Neuron 28:617–625.

Reale RA, Calvert GA, Thesen T, Jenison RL, Kawasaki H, Oya H, Howard MA, Brugge JF (2007): Auditory-visual processing represented in the human superior temporal gyrus. Neuroscience 145:162–184.

Ridderinkhof KR, Ullsperger M, Crone EA, Nieuwenhuis S (2004): The role of the medial frontal cortex in cognitive control. Science 306:443–447.

Rizzolatti G, Arbib MA (1998): Language within our grasp. Trends Neurosci 21:188–194.

Ross LA, Saint-Amour D, Leavitt VM, Javitt DC, Foxe JJ (2007): Do you see what I am saying? Exploring visual enhancement of speech comprehension in noisy environments. Cereb Cortex 17:1147–1153.

Sekiyama K, Kanno I, Miura S, Sugita Y (2003): Auditory-visual speech perception examined by fMRI and PET. Neurosci Res 47:277–287.

Sekuler R, Sekuler AB, Lau L (1997): Sounds alter visual motion perception. Nature 385:308.

Semple MN, Scott BH (2003): Cortical mechanisms in hearing. Curr Opin Neurobiol 13:167–173.

Skipper JI, Nusbaum HC, Small SL (2005): Listening to talking faces: Motor cortical activation during speech perception. Neuroimage 25:76–89.

Skipper JI, van Wassenhove V, Nusbaum HC, Small SL (2007): Hearing lips and seeing voices: How cortical areas supporting speech production mediate audiovisual speech perception. Cereb Cortex 17:2387–2399.

Stein BE, Meredith MA, Huneycutt WS, McDade L (1989): Behavioural indices of multisensory integration: Orientation to visual cues is affected by auditory stimuli. J Cogn Neurosci 1:12–24.

Sumby WH, Pollack I (1954): Visual contribution to speech intelligibility in noise. J Acoust Soc Am 26:212–215.

Szycik GR, Tausche P, Münte TF (2007): A novel approach to study audiovisual integration in speech perception: Localizer fMRI and sparse sampling. Brain Res. [Epub ahead of print].

Talairach J, Tornoux P (1988): Co-Planar Stereotactic Atlas of the Human Brain. Stuttgart: Thieme.

Tokimura H, Tokimura Y, Oliviero A, Asakura T, Rothwell JC (1996): Speech-induced changes in corticospinal excitability. Ann Neurol 40:628–634.

Tomaiuolo F, MacDonald JD, Caramanos Z, Posner G, Chiavaras M, Evans AC, Petrides M (1999): Morphology, morphometry and probability mapping of the pars opercularis of the inferior frontal gyrus: An in vivo MRI analysis. Eur J Neurosci 11: 3033–3046.

van Atteveldt N, Formisano E, Goebel R, Blomert L (2004): Integration of letters and speech sounds in the human brain. Neuron 43:271–282.

Wright TM, Pelphrey KA, Allison T, McKeown MJ, McCarthy G (2003): Polysensory interactions along lateral temporal regions evoked by audiovisual speech. Cereb Cortex 13:1034–1043.