Brain activation during audiovisual exposure anticipates future perception of ambiguous speech

NeuroImage 57 (2011) 1601–1607

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/ynimg

Brain activation during audiovisual exposure anticipates future perception of ambiguous speech

Niclas Kilian-Hütten [a,b,*], Jean Vroomen [c], Elia Formisano [a,b]

[a] Dept. of Cognitive Neuroscience, Faculty of Psychology and Neuroscience, Maastricht University, The Netherlands
[b] Maastricht Brain Imaging Center (MBIC), Maastricht, The Netherlands
[c] Dept. of Psychology, Tilburg University, The Netherlands

ARTICLE INFO

Article history:
Received 9 February 2011
Revised 21 April 2011
Accepted 16 May 2011
Available online 25 May 2011

Keywords:
fMRI
Audiovisual
McGurk
Auditory perception
Cross-modal recalibration
Implicit learning

ABSTRACT

In modern perceptual neuroscience, the focus of interest has shifted from a restriction to individual modalities to an acknowledgement of the importance of multisensory processing. One particularly well-known example of cross-modal interaction is the McGurk illusion. It has been shown that this illusion can be modified, such that it creates an auditory perceptual bias that lasts beyond the duration of audiovisual stimulation, a process referred to as cross-modal recalibration (Bertelson et al., 2003). Recently, we have suggested that this perceptual bias is stored in auditory cortex, by demonstrating the feasibility of retrieving the subjective perceptual interpretation of recalibrated ambiguous phonemes from functional magnetic resonance imaging (fMRI) measurements in these regions (Kilian-Hütten et al., 2011). However, this does not explain which brain areas integrate the information from the two senses and represent the origin of the auditory perceptual bias. Here we analyzed fMRI data from audiovisual recalibration blocks, utilizing behavioral data from perceptual classifications of ambiguous auditory phonemes that followed these blocks later in time. Adhering to this logic, we could identify a network of brain areas (bilateral inferior parietal lobe [IPL], inferior frontal sulcus [IFS], and posterior middle temporal gyrus [MTG]), whose activation during audiovisual exposure anticipated auditory perceptual tendencies later in time. We propose a model in which a higher-order network, including IPL and IFS, accommodates audiovisual integrative learning processes, which are responsible for the installation of a perceptual bias in auditory regions. This bias then determines constructive perceptual processing.

Introduction

Traditionally, research on human perception has focused on unimodal information processing. However, in real life we are constantly confronted with multimodal input, i.e. information coming from several/all senses. Importantly, multimodal processing is not restricted to the mere addition of a second channel. Instead, these channels interact, affecting processing in individual modalities. For instance, in face-to-face spoken conversation, lip reading provides for vital information, enhancing auditory understanding. This is especially true in suboptimal acoustic settings. Furthermore, research has consistently demonstrated intermodal effects on behavior (Shimojo and Shams, 2001). Congruent multisensory input typically lowers detection thresholds (Frassinetti et al., 2002), shortens reaction times (Forster et al., 2002; Schroger and Widmann, 1998), and decreases saccadic eye movement latencies (Hughes et al., 1994). Incongruent multisensory input has opposite consequences (Sekuler et al., 1997).

One particularly interesting effect of audiovisual interaction is the McGurk illusion (McGurk and MacDonald, 1976). Here the addition of the visual modality does not only affect the quality of the auditory percept, but it actively alters its identity. Typically, an auditory disyllable (e.g., /aba/) uttered by a speaker is presented simultaneously with a video showing the speaker's lips pronouncing an incongruent visual disyllable (/aga/). The integration of this bimodal input results in a percept that is intermediate between the auditory and the visual channels (/ada/). Neuroimaging studies investigating the McGurk illusion have identified networks very similar to the ones found with other audiovisual paradigms (Beauchamp et al., 2010; Benoit et al., 2010; Fingelkurts et al., 2003; Hertrich et al., 2011; Jones and Callan, 2003; Jones and Jarick, 2006; Mottonen et al., 2002; Sekiyama et al., 2003; Wiersinga Post et al., 2010).

It has been shown that the McGurk illusion can be modified in such a way that the biasing perceptual effect lasts beyond the audiovisual stimulation itself (Bertelson et al., 2003). The decisive factor for this is the use of an ambiguous auditory component (A?, between /aBa/ and /aDa/). Exposure to such an ambiguous stimulus

* Corresponding author at: P.O. Box 616, 6200 MD Maastricht, The Netherlands. Fax: +31 43 388 4125.
E-mail addresses: niclas.kilian-hutten@maastrichtuniversity.nl (N. Kilian-Hütten), j.vroomen@uvt.nl (J. Vroomen), e.formisano@maastrichtuniversity.nl (E. Formisano).

1053-8119/$ – see front matter © 2011 Elsevier Inc. All rights reserved.
doi:10.1016/j.neuroimage.2011.05.043

dubbed onto a video of a face pronouncing /aBa/ or /aDa/ (A?Vb and A?Vd, respectively) selectively increases the proportion of corresponding responses in subsequent audio-only forced-choice trials, thus eliciting an after effect ("cross-modal recalibration"). In the original study, an audiovisual exposure block of 8 adapters (i.e., identical videos) sufficed to significantly bias the perceptual categorization of 6 subsequent auditory post-tests.

Recently, we have demonstrated that it is possible to retrieve the perceptual interpretation of the ambiguous phonemes presented in these post-tests from functional magnetic resonance imaging (fMRI) measurements of brain activity in auditory areas in the superior temporal cortex, most prominently on the posterior bank of the left Heschl's gyrus and sulcus and in the adjoining left planum temporale (Kilian-Hütten et al., 2011). We have proposed that the perceptual bias, which is responsible for the behavioral effect, is installed by the cross-modal recalibration mechanism and is stored within these regions. However, this leaves open the question where the top-down audiovisual recalibration effect originates, i.e. which networks are responsible for prejudicing auditory cortex in creating perceptual interpretations.

The goal of the present study was to investigate the neural origin of cross-modal recalibration, employing McGurk after effects. We analyzed fMRI measurements from the same data set as used by Kilian-Hütten et al. (2011). However, instead of looking at data from the auditory post-tests, here we concentrated on the audiovisual recalibration blocks. We expected a simple comparison between the activation elicited by the audiovisual stimuli and baseline to replicate the results from prior studies on audiovisual processing and the McGurk illusion and to identify a set of areas similar to the ones found in these studies. However, our recalibration design biases perception beyond the time of audiovisual stimulation, and thus provided us with the unique possibility to go beyond the mere identification of active areas. In order to determine which subset of active areas drives the lasting biasing effect on auditory perception, we applied a behaviorally defined contrast. The strength of the recalibration effect is variable from one particular exposure block to another and can be quantified in terms of the number of auditory post-tests perceived in line with the type of exposure block (A?Vb or A?Vd). We could then use these values to identify brain regions whose activation during the exposure blocks varied with the strength of the recalibration effect. The hemodynamic response in these areas, thus, 'predicted' the perceptual tendency in a separate, later time window. Hence, this approach provides the possibility to go beyond the basic identification of responsive regions in an audiovisual paradigm and allows identifying those regions which are responsive to audiovisual stimulation and which, further, are likely functionally relevant in driving the effect of these stimuli in biasing subsequent auditory perception. In order to further investigate whether the ROIs identified with the behaviorally weighted contrast exert a task-related influence on auditory regions, which are involved in the perceptual interpretation of the auditory post-tests (Kilian-Hütten et al., 2011), we carried out a psychophysiological interaction analysis (PPI; Friston et al., 1997). PPI is a method to investigate how the influence of the activity in one region of the brain on the activity of spatially distant regions changes depending on the task context.

Methods

Participants

Twelve healthy native Dutch students of the University Maastricht (5 male, mean age 24.83) were recruited to participate in the study. All but one participant were right-handed. None of the participants had a history of hearing loss or neurological abnormalities. Approval for the study was granted by the Ethical Committee of the Faculty of Psychology at the University of Maastricht.

Stimuli

Based on the psychophysical study by Bertelson et al. (2003), the stimulation entailed digital auditory and visual recordings of a male Dutch speaker pronouncing the syllables /aBa/ and /aDa/. The two auditory stimuli were 640-ms stimuli with 240-ms stop closure. From these, a place-of-articulation continuum was synthesized by means of varying the F2 formant by equal steps of 39 Mel, resulting in 9 different stimuli, ranging from a clear /aba/ via seven ambiguous stimuli to a clear /ada/. The audiovisual stimuli were synthesized by pairing the visual recordings of the speaker pronouncing /ada/ and /aba/, respectively, with the most ambiguous auditory stimulus, determined as such in a pre-test (please see the online supplementary material for example stimuli).

Behavioral pre-test

In an auditory pre-test outside the fMRI scanner the most ambiguous auditory stimulus from the /aBa/-to-/aDa/ continuum was determined for each participant individually (Bertelson et al., 2003). All stimuli on the continuum were presented (98 trials in total). The presentation frequency was biased, so that the five central stimuli were presented on 14 trials, the second and tenth stimuli were presented on 8 trials, and the first and the eleventh stimuli were presented on 6 trials. Stimuli were presented binaurally through loudspeakers. Participants indicated whether they had perceived /ada/ or /aba/ by means of a button press. We estimated each participant's most ambiguous auditory token (*A?*), which was then used in the actual experiment.

Scanning parameters

We collected functional brain data on a 3-Tesla fMRI scanner (head set-up, Siemens Allegra, Siemens, Erlangen, Germany) at the Maastricht Brain Imaging Center (M-BIC) in Maastricht, The Netherlands. For each participant, 2 functional runs of 665 volumes were acquired. A high-resolution structural scan (voxel size $= 1 \times 1 \times 1$ mm^3) was collected using a T1-weighted 3D ADNI sequence (TR = 2050 ms, TE = 2.6 ms, 192 sagittal slices) for later overlay of the functional data. All scans were carried out in a single session for each participant.

For functional images, a BOLD-sensitive echo-planar imaging (EPI) sequence was used (matrix 64×64, 27 slices, slice thickness 3 mm, field of view, FOV, 192×192 mm^3, resulting voxel size $3 \times 3 \times 3$ mm^3, TE/TR slice 30/55.5, FA = 90°). Volume acquisition was clustered in the beginning of each TR, leaving a silent delay within each TR, during which stimuli were presented in the absence of EPI noise. This was done in order to optimize stimulus audibility, an approach which has been shown to be highly efficient in auditory fMRI paradigms (Jancke et al., 2002; van Atteveldt et al., 2004, 2007). Hence, the effective TR was 2000 ms, including 1500 ms of sequence scanning time and a 500 ms silent delay. Stimuli were presented and synchronized with the MR pulses using the software package "Presentation" (Neurobehavioral Systems).

Experimental procedure

Each functional run consisted of 10 mini runs, which were composed of bimodal exposure blocks and slow event-related auditory post-tests. The bimodal (recalibration) blocks entailed the repeated presentation of videos, which consisted of the participant's A? paired with the visual recording of the speaker's lips pronouncing either /aba/ (Vb) or /ada/ (Vd). Each block was composed of 8 identical videos (block A?Vb or block A?Vd). The interstimulus interval (ISI) was one TR (=2000 ms). For each run, 5 A?Vb and 5 A?Vd blocks were presented in random order (160 trials in total). In order to warrant attentional focus, participants were instructed to press a button

whenever a small white dot (12 pixels) appeared on the speaker's upper lip. This was the case once per block at a random position.

Each bimodal recalibration block was followed by 6 auditory-only post-tests (total: 120 trials), which consisted of forced-choice /aba/-/ada/ judgments. The *A?* token and the two tokens closest to it on the continuum were presented, twice each. Owing to the use of a slow event-related design, the jittered ISI was 6 TR (=12 s) on average (Fig. 1).

Data pre-processing

Functional and anatomical images were analyzed using BrainVoyager QX (Brain Innovation, Maastricht, The Netherlands). Several pre-processing steps were performed: sinc interpolated slice-time correction, 3D-motion correction to correct for common small head movements by spatially aligning all volumes to the first volume by rigid body transformations, linear trend removal, and temporal high-pass filtering to remove low-frequency nonlinear drifts of 7 or less cycles per time course. Functional slices were then co-registered to the structural volume on the basis of positioning parameters from the scanner and manual adjustments to ensure optimal fit. Subsequently, they were transformed into Talairach Space. All individual brains were segmented at the gray/white matter boundary using a semi-automatic procedure based on intensity values implemented in BrainVoyager QX. For all group analyses, cortex-based alignment was used to assure optimized spatial matching of cortical locations (i.e., vertices) between participants (Goebel et al., 2006).

Statistical analysis

Functional runs were analyzed using voxel-wise multiple linear regression (GLM) of the BOLD response time course. All analyses were performed at single-subject and group levels and all experimental conditions were modeled as predictors. In all GLMs, predictor time courses were convolved with a hemodynamic response function (where both the BOLD response and undershoot are modeled by a gamma function) to adjust for the hemodynamic delay (Friston et al., 1998).

Linear contrasts served to implement comparisons of interest. To identify brain regions involved in the processing of the audiovisual (recalibration block) stimuli, a contrast of audiovisual stimulation (all recalibration blocks) versus baseline was applied. In order to find a subset of these areas whose activation during the recalibration blocks would predict post-test performance / perception, a contrast based on behavioral performance was then utilized. To achieve this, we implemented a predictor, in which each recalibration block was coded with an individual, behaviorally defined value. These values were obtained by identifying the number of responses in line with recalibration from the six auditory post-tests belonging to a specific block (i.e., /ada/ after a A?Vd block and /aba/ after an A?Vb block). These values between zero and six were then normalized across all blocks of one type for a given participant and orthogonalized with respect to the overall activation contrast. We then applied this behaviorally defined contrast in order to identify ROIs whose activation pattern was correlated with post-test performance. ROIs exhibiting the strongest significance for this contrast, thus, demonstrated a BOLD response pattern, whose modulation on a block level best predicted *subsequent* behavioral / perceptual performance.

These statistical analyses were performed following a two-step procedure; first, contrasts were implemented on the individual level. These contrasts were then entered into a second-level, random-effects model in order to allow for inferences at the population level (Friston et al., 1999). In order to correct for false positives, a cluster-size threshold procedure was employed (Forman et al., 1995; Goebel et al., 2006). For this, cluster-level false-positive rates were estimated iteratively (500 iterations; Monte Carlo simulation). The minimum cluster-size threshold (based on its observed relative frequency) which yielded a cluster-level false-positive rate (alpha) of 5% was then applied to the statistical maps.

Psychophysiological interaction

For the PPI analysis (Friston et al., 1997), we identified a seed region in the left auditory cortex (lAC) individually in all twelve subjects based on the [audiovisual blocks>baseline] contrast (see Table 1). This was done following Kilian-Hütten et al. (2011), who identified a distributed pattern of voxels that encoded the perceptual interpretation of the auditory post-test and was located predominantly in left auditory cortex. We then investigated the functional connectivity of the remaining voxels with lAC during the audiovisual exposure blocks compared to baseline. This was done by employing a GLM in which we implemented a predictor for the activity time course in the seed region and one for the interaction between the task (audiovisual blocks) and the time course of the seed (PPI). This latter predictor, thus, served to identify regions which exhibited increased functional connectivity with the seed region during the audiovisual blocks relative to baseline. We also included all other predictors (auditory post-tests) in the analysis, in order to account for common effects and only identify those effects that were unique to the PPI. Group maps were created be running this analysis as a random-effects GLM on the cortex-based aligned data and thresholding the results using a cluster-size thresholding approach such as described above. Additionally, the same analyses were implemented as random-effects ROI GLMs in the ROIs identified by the behaviorally weighted contrast in order to investigate whether these regions exhibit increased connectivity during the recalibration blocks.

Results

Behavioral results

Individual ambiguous sounds chosen on the basis of the pre-tests were item 4 for eight participants and item 5 for the four others. Exposure to the audiovisual recalibration blocks significantly influenced the perception of the auditory post-test stimuli (see Fig. 2 and Kilian-Hütten et al., 2011). An analysis of variance (ANOVA) over all sounds (A?, A?d, A?b) demonstrated a highly significant difference in proportion of /aBa/ responses between sounds that were preceded by an A?Vd block versus an A?Vb block (.53 following A?Vb versus .31 following A?Vd, $F = 75.682$, $p < .001$).

This significant difference could be shown for all three sounds separately (***A?***: .60 following A?Vb versus .25 following A?Vd,

Fig. 1. Schematic overview of one example mini run (A?Vb). Each run consisted of 10 such mini runs, which each entailed a bimodal exposure block (A?Vd or A?Vb) and 6 auditory post-test trials. The bimodal blocks were presented within 8 TRs, each of which entailed 1500 ms of scanning and 500 ms of video presentation. Between auditory post-test trials, fixation periods averaged 6 TRs.

Table 1
Seed regions for the PPI analysis. ROIs were identified individually in each subject on the basis of overall audiovisual block activation. The average number of voxels per ROI was 443 (SD = 134).

Subject	Talairach (Center of gravity)		
	x	y	z
AB	−38	−34	15
AF	−53	−26	14
CS	−43	−29	10
JR	−44	−23	5
KJ	−39	−31	14
MC	−51	−24	5
MM	−37	−33	15
NS	−44	−27	10
NSt	−43	−21	8
RG	−49	−23	6
RV	−39	−32	12
TG	−43	−30	5

F = 68.692, $p < .001$; ***A?b:*** .85 following A?Vb versus .64 following A?Vd, F = 31.546, $p < .001$; ***A?d:*** .15 following A?Vb versus .05 following A?Vd, F = 12.757, $p < .001$). These results demonstrate the effectiveness of the cross-modal recalibration design in biasing subsequent auditory perception and replicate the results from Bertelson et al. (2003).

fMRI results

The first contrast of interest was a comparison of activity in response to the audiovisual blocks versus baseline ([AV>baseline]). Fig. 3 shows an overview of brain areas active in response to this audiovisual stimulation. Among these are early unimodal areas (primary and extrastriate visual areas and primary and early auditory areas), as well as areas which have been suggested to be involved in audiovisual integration (STG/STS, inferior parietal lobe and insula cortex), and higher-order areas, possibly involved in attention, action understanding and intellectualization processes (inferior frontal gyrus [IFG], inferior frontal sulcus [IFS] and premotor cortex).

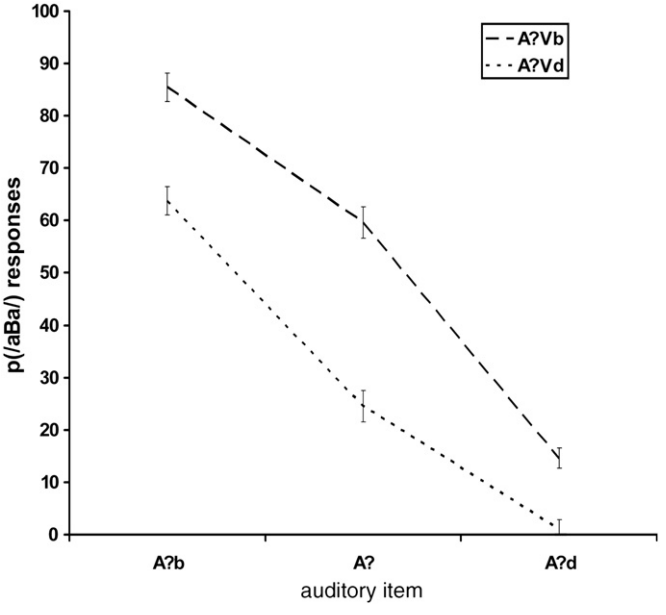

Fig. 2. Behavioral results of the auditory post-tests. For the participant's ambiguous auditory item and its two neighbors on the continuum, the graph shows the proportion of /aba/ classifications after exposure to an audiovisual adapter comprised of the ambiguous item paired with either visual /aba/ (A?Vb) or with visual /ada/ (A?Vd). For all three auditory items, the difference in proportion /aba/ responses after exposure to the A?Vb adapter vs. the A?Vd adapter was significant.

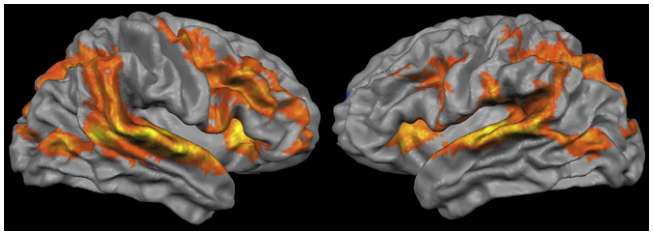

Fig. 3. Group results for the random-effects GLM and the contrast [AV>baseline] overlaid on the average hemispheres obtained from the cortex-based alignment procedure (cluster-size corrected to include only clusters of 168 mm^2 or more, corresponding to a cluster-level false-positive rate (alpha) of 5%). Shown are early auditory and visual areas, STG/STS, insula, IFS and IFG, IPL and premotor cortex.

The activation pattern is very symmetrical with respect to hemisphere, except for the frontal activation, which seems to be stronger in the right hemisphere.

The activation map obtained from the behaviorally weighted contrast is shown in Fig. 4 superimposed onto the average hemispheres obtained from the cortex-based alignment procedure. Activation peaks are listed in Table 2 along with their Talairach coordinates and their z and p values for the behaviorally weighted contrast.

This analysis highlights very symmetric areas of robust activation in the right and left inferior parietal lobe (IPL), approximately corresponding to Brodmann's area (BA) 40. Additionally, we found ROIs in the inferior frontal sulcus (IFS, BA 44) and the posterior middle temporal gyrus (MTG, BA 21), in both hemispheres.

The relation between activation level and strength of the recalibration effect is further illustrated in Fig. 5. Percent BOLD signal changes in blocks that strongly recalibrated later perception (4–6 auditory post-tests in line with the direction of recalibration) were higher than in those blocks that did not (0–3 auditory post-tests in line with the direction of recalibration) for the six ROIs, identified in individual participants.

PPI results

The group map from the PPI analysis is shown in Fig. 6. Although this map includes a broader network than the one obtained with the behaviorally weighted contrast, it bears quite a clear resemblance with it, including bilateral IPL, IFS and MTG (although there is no perfect overlap between ROIs). Additionally, it includes dorsal and ventral stream visual areas in the occipital lobe. The ROI GLM analyses corroborate the impression of resemblance and demonstrate a significant increase of connectivity during the recalibration blocks for all ROIs obtained with the behaviorally weighted contrast, except for left IFS, which just did not reach significance (Table 3).

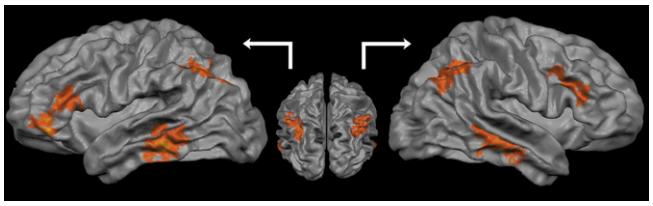

Fig. 4. Group results for the behaviorally weighted contrast (random-effects GLM of cortex-based aligned data) overlaid on the average hemispheres obtained from the cortex-based alignment procedure (cluster-size corrected to include only clusters of 174 mm^2 or more, corresponding to a cluster-level false-positive rate (alpha) of 5%). Shown are bilateral IPL, IFS, and posterior middle temporal gyrus.

Table 2
ROIs obtained from the behaviorally weighted contrast in the cortex-based aligned group analysis, along with the spatially closest Brodmann area, Talairach coordinates (center of gravity) and the associated z and p values for the behaviorally weighted contrast (random-effects GLM).

Brain region	Brodmann area	Side	Talairach (center of gravity)			Behaviorally weighted contrast	
			x	y	z	z	p
Inferior parietal lobe	39/40	Right	41	−59	36	2.627	<.01
Inferior parietal lobe	39/40	Left	−38	−60	36	2.916	<.01
Middle temporal gyrus	21	Right	57	−29	−6	2.935	<.01
Middle temporal gyrus	21	Left	−55	−34	−8	2.921	<.01
Inferior frontal sulcus	09/46	Right	42	12	30	2.390	<.05
Inferior frontal sulcus	09/46	Left	−43	32	12	2.834	<.01

Discussion

The goal of the present study was to investigate the neural mechanisms underlying audiovisual recalibration of auditory phoneme perception. In order to identify brain regions which were involved in the processing of the audiovisual stimuli employed in this study, we contrasted blocks of videos against baseline. A broad, rather symmetric network of brain areas showed a significant increase in BOLD response during these blocks. ROIs included primary and extrastriate visual areas, primary and early auditory areas, STG/STS, inferior frontal sulcus (IFS), premotor cortex, and inferior parietal lobe, spreading out to touch angular and supermarginal gyri and the intraparietal sulcus (jointly referred to as IPL from now on). These results replicate earlier neuroimaging studies on cross-modal processing and the McGurk effect (Beauchamp et al., 2010; Benoit et al., 2010; Calvert, 2001; Calvert and Campbell, 2003; Calvert et al., 2000; Calvert and Thesen, 2004; Fingelkurts et al., 2003; Hertrich et al., 2011; Jones and Callan, 2003; Jones and Jarick, 2006; Meienbrock et al., 2007; Mottonen et al., 2002; Nishitani and Hari, 2002; Ojanen et al., 2005; Olson et al., 2002; Saito et al., 2005; Sekiyama et al., 2003; Wiersinga Post et al., 2010; Wright et al., 2003). The activation in early unimodal regions (due to auditory and visual stimulation, respectively) and in premotor cortex (related to the button press response) is in line with expectations, as is the activation in the superior temporal lobe. These regions (STS/STG) have consistently been found to be implicated in the integration of audiovisual, and more specifically, face-voice stimuli (Calvert, 2001; Calvert and Campbell, 2003; Calvert et al., 2000; Calvert and Thesen, 2004; Jancke et al., 2002; Olson et al., 2002; Sekiyama et al., 2003; Wright et al., 2003). At least two recent studies corroborate this view by demonstrating that these regions follow the principle of inverse effectiveness in humans (Stevenson and James, 2009; Werner and Noppeney, 2009) and one study demonstrates that transcranial magnetic stimulation over STS disrupts the perception of the McGurk illusion (Beauchamp et al., 2010). Although we did not specifically test for this, our results are compatible with the STS/STG area's suggested role in the integration of audiovisual stimuli.

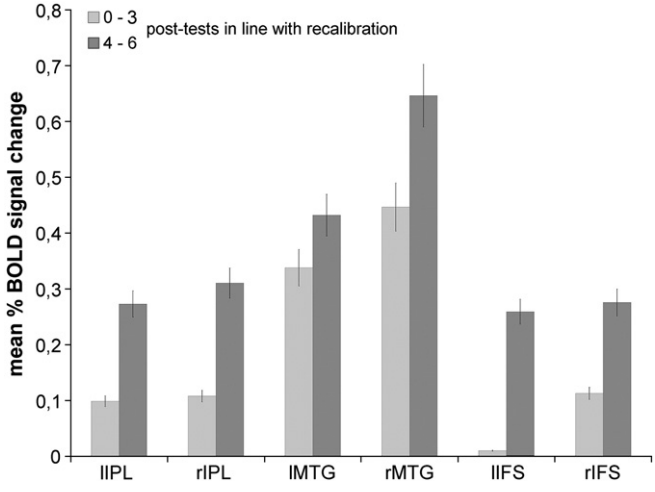

Fig. 5. BOLD activation is shown for weak versus strong recalibration blocks in several ROIs (IPL, IFS and MTG; in the left and right hemisphere, respectively). Regions were identified individually in each participant on the basis of the behaviorally weighted GLM contrast. Shown is the mean percentage of BOLD signal change in blocks that strongly recalibrated later auditory perception (4–6 post-tests in line with the direction of recalibration) versus those that did not (0–3 post-tests in line with the direction of recalibration). Error bars denote ± 1 standard error.

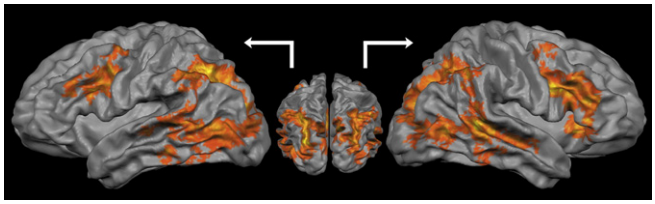

Fig. 6. Group results for the psychophysiological interaction analysis (random-effects GLM of cortex-based aligned data) overlaid on the average hemispheres obtained from the cortex-based alignment procedure (cluster-size corrected to include only clusters of 430 mm^2 or more, corresponding to a cluster-level false-positive rate (alpha) of 5%). Shown are bilateral IPL, IFS, posterior middle temporal gyrus, and dorsal and ventral stream visual areas in the occipital lobe.

The remaining regions are located in higher-order association cortex, namely IPL and IFS. These are the areas that seem to be functionally relevant in driving the effect of audiovisual recalibration of auditory phoneme perception. In order to identify the subset of active regions for which this is true, we made use of participants' online behavior (i.e., the efficiency of the individual blocks of audiovisual stimulation in eliciting the recalibration effect). This allowed identifying those areas whose hemodynamic response profile predicted the strength of the recalibration effect on a block-to-block basis. Since this reasoning relies on a correlation between hemodynamic response and participants' behavior, it does not permit the formulation of definite causal claims. However, the fact that the modulation of the hemodynamic response in these areas considerably preceded the auditory stimuli and forced-choice perceptual judgments in time, does strongly suggest a functional role in the audiovisual recalibration of auditory phoneme perception. This is further corroborated by the results from the psychophysiological interaction analysis, which show that this network of areas exhibits increased functional/effective connectivity with the left auditory cortex during recalibration blocks relative to baseline. It has to be noted that the present paradigm does not lend itself easily to more complex connectivity analysis approaches such as dynamic causal modeling or Granger causality mapping. This is mainly due to the fact that the encoding of the perceptual bias is represented in distributed patterns (which we showed before using multivoxel pattern analysis

Table 3
Results from the psychophysiological interaction ROI GLMs for ROIs obtained from the behaviorally weighted contrast. Z and p values are shown.

Brain region	Side	Psychophysiological interaction	
		z	p
Inferior parietal lobe	Right	3.146	<.01
Inferior parietal lobe	Left	3.814	<.001
Middle temporal gyrus	Right	2.762	<.01
Middle temporal gyrus	Left	2.672	<.01
Inferior frontal sulcus	Right	3.268	<.01
Inferior frontal sulcus	Left	1.949	<.052

(Kilian-Hütten et al., 2011)), while the ROIs in the present analysis were identified with a univariate approach. We are not aware of any past demonstrations of complex connectivity analyses in this situation and have, hence, restricted ourselves to the more straightforward PPI analysis. It is a crucial challenge for future research to devise functional and effective connectivity analysis techniques that integrate univariate and multivariate results.

The following section explores our interpretation of what the functional role of the identified network might be. Subsequently, alternative explanations will be discussed.

Integrative learning effects on constructive perceptual processes

The very symmetrical ROIs in the IPL are the most robust regions obtained from the behaviorally weighted contrast. One possible functional explanation relates to the brain's pervasive ability and need to categorize incoming information. A recent study (Raizada and Poldrack, 2007) investigated the selective amplification of phonetic stimulus differences. Stimuli from a phonetic continuum were presented in pairs. In some brain regions the activation for pairs which crossed a category boundary was amplified, while activation in response to pairs which did not to cross the boundary was suppressed. It could be shown that activation in the left supramarginal gyrus most strongly reflected such categorization. This suggests a role for the supramarginal gyrus (which is part of our ROI in the IPL) in constructive perceptual processes and is in line with the present study. It can be speculated that the presentation of the audiovisual stimuli provides information that is used by the supramarginal gyrus in order to then influence regions earlier in the hierarchy in building the percept, i.e. in categorizing the auditory stimuli. Interestingly, another study investigating phonetic invariance observed a region in the left IFS which was insensitive to acoustic changes within a phonetic category while demonstrating sensitivity to changes between phonetic categories (Myers et al., 2009). We also found ROIs in this region bilaterally for the behaviorally weighted contrast.

A recent study (Naumer et al., 2009) may expand this general view. In order to study the effect of learning associations between novel sounds and images on audiovisual object representation, the authors compared the integration of such stimuli before and after training of specific pairings. Intriguingly, the network of regions they found for post-training integration is very similar to the network we found here. More specifically, while pre-training activation was largely confined to the right inferior frontal cortex (coinciding with our ROI in the IFS), this activation became bilateral after training. Additionally, activation was observed bilaterally in the inferior parietal lobe, seemingly in agreement with the location of our IPL ROIs. Even our ROIs in posterior middle temporal gyrus can be found back in these data, namely when comparing post-training activation in response to learned associations versus mismatched audiovisual stimuli. The conspicuous agreement of this network with the network of regions we found with our behaviorally weighted contrast suggests that there should be a role of these areas that is shared by both paradigms. What the paradigms have in common is learning of an audiovisual association. While learning is explicitly required in the study by Naumer and colleagues (2009), it is of implicit nature in the present paradigm. This fits the definition put forward by Gilbert et al. (2001, p. 681), who refer to perceptual learning as "improving one's ability, with practice, to discriminate differences in the attributes of simple stimuli". Improvement has to be viewed in relative and contextual terms here. In the case of recalibration, the disambiguating information from audiovisual exposure biases auditory perception such that it can be regarded as improved in reference to the (momentary) demands of sensory reality. It, thus, qualifies as a case of perceptual learning. Cross-modal learning, in this case, does not only induce plastic changes of the neural substrates of audiovisual processing, but these also have a functional role. This is constructive in nature. Every time an association between the ambiguous auditory stimuli and a certain lip movement is learned, this updates our interpretation of the world, i.e. biases subsequent perception. This is the same role that can be assigned to the supramarginal gyrus on the basis of the study by Raizada and Poldrack (2007). The major difference lies only in the information that is used as a basis for the constructive process. While in our paradigm it is associations with lip movements, in their study it is the contrast with phonemes from the other side of the category boundary that forms this basis. Extreme cases of such multisensory plastic functional changes are apparent in phenomena such as acquired synesthesia (Beauchamp and Ro, 2008; Naumer and van den Bosch, 2009).

Alternative explanations: verbal working memory and cognitive control

There are some alternative interpretations of our data which demand some consideration. Both inferior parietal and inferior frontal cortices have been linked to different verbal working memory processes (Baldo and Dronkers, 2006; Hertrich et al., 2011). It is conceivable that stronger activation of such a verbal working memory network (possibly due to block-to-block variability of attention) would reflect better retention of the phonemes presented during the audiovisual blocks, which then would exert a bigger influence on the perceptual interpretation of the auditory stimuli presented afterwards. If this was the decisive factor behind audiovisual recalibration, it would be hard to assess whether this effect was really due to actual constructive perceptual changes. Alternatively, a more cognitive response bias could have an impact here. Support for this notion comes from a meta analysis that linked activation in a region within the IFS to cognitive control in a number of switching paradigms (Derrfuss et al., 2005). In such studies (task-switching, set-shifting, stimulus-response reversal studies, and color-word Stroop studies) the participant is required to select between two competing responses. These studies would suggest then that part of the recalibration effect was due to a response bias as opposed to a pure perceptual shift. In view of the automaticity of the immediate effect that the lip movement information exerts on auditory perception during audiovisual exposure and the strength of the recalibration effect, we do, however, consider this possibility unlikely. Evidence in support of this view is threefold. First, when no ambiguous, but an unambiguous auditory component is used, instead of recalibration effects, selective speech adaptation is found (Bertelson et al., 2003; Vroomen et al., 2007). In other words, only pairing an ambiguous auditory component with a disambiguating lip movement increases the probability of similar percepts later in time. Stimuli that are unambiguous in both modalities have the opposite effect, i.e. they decrease the probability of similar percepts. This discrepancy is not explainable on the basis of memory-related response biases, which should be the same in both cases. Secondly, engaging participants in a visuospatial or verbal working memory task under

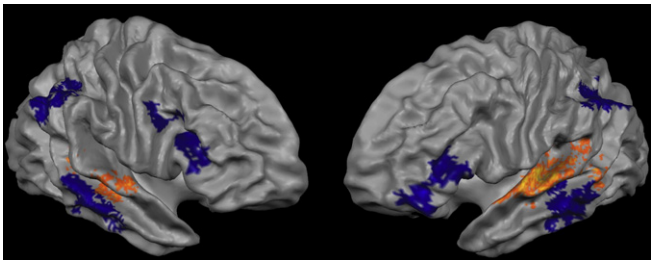

Fig. 7. The proposed model. A higher-order network (in blue) including IPL, IFS, and MTG is suggested to process integrative learning effects, and consequently install a perceptual bias in auditory regions (in red and orange), most prominently the left Heschl's sulcus and the planum temporale, influencing future constructive auditory perception.

different memory load conditions does not affect phonetic recalibration (Baart and Vroomen, 2010), which would be expected in case of a working memory-based response bias. Thirdly, we have shown that it is possible to predict the perceptual interpretation of the auditory post-tests on a single-trial basis from activation in the auditory cortex (Kilian-Hütten et al., 2011). Again, this speaks against a cognitive response bias and suggests a perceptual effect, which is stored in hierarchically lower regions.

Conclusion

We showed that ambiguous sounds paired with lip movements, which effectively disambiguated the sounds, strongly influenced the dichotic categorization of these ambiguous sounds presented in isolation later in time. The hemodynamic response during audiovisual exposure within a network of brain areas, namely bilateral and very symmetric regions in the IPL, IFS, and posterior middle temporal gyrus, predicted later speech perception. Drawing upon published empirical work, we interpret the activation in this network to reflect integrative audiovisual learning effects, which in turn affect constructive (auditory) perceptual processes.

Elsewhere (Kilian-Hütten et al., 2011), we have provided evidence that suggests that these constructive perceptual processes themselves are located in hierarchically lower, auditory regions (the posterior bank of the left Heschl's gyrus and sulcus and the adjoining left planum temporale). Thus, together with the results from this article, a complete model of cross-modal recalibration emerges, in which a higher-order network of brain areas (including IPL and IFS) is involved in audiovisual integrative learning processes, which lead to the installation of a perceptual bias in auditory regions (Fig. 7). This bias in turn determines constructive perceptual processes.

Supplementary materials related to this article can be found online at doi:10.1016/j.neuroimage.2011.05.043.

Acknowledgments

We would like to thank Fren Smulders, Martin Frost, and Giancarlo Valente for valuable discussions.

References

Baart, M., Vroomen, J., 2010. Phonetic recalibration does not depend on working memory. Exp. Brain Res. 203, 575–582.
Baldo, J.V., Dronkers, N.F., 2006. The role of inferior parietal and inferior frontal cortex in working memory. Neuropsychology 20, 529–538.
Beauchamp, M.S., Ro, T., 2008. Neural substrates of sound-touch synesthesia after a thalamic lesion. J. Neurosci. 28, 13696–13702.
Beauchamp, M.S., Nath, A.R., Pasalar, S., 2010. fMRI-guided transcranial magnetic stimulation reveals that the superior temporal sulcus is a cortical locus of the McGurk effect. J. Neurosci. 30, 2414–2417.
Benoit, M.M., Raij, T., Lin, F.H., Jaaskelainen, I.P., Stufflebeam, S., 2010. Primary and multisensory cortical activity is correlated with audiovisual percepts. Hum. Brain Mapp. 31, 526–538.
Bertelson, P., Vroomen, J., De Gelder, B., 2003. Visual recalibration of auditory speech identification: a McGurk aftereffect. Psychol. Sci. 14, 592–597.
Calvert, G.A., 2001. Crossmodal processing in the human brain: insights from functional neuroimaging studies. Cereb. Cortex 11, 1110–1123.
Calvert, G.A., Campbell, R., 2003. Reading speech from still and moving faces: the neural substrates of visible speech. J. Cogn. Neurosci. 15, 57–70.
Calvert, G.A., Thesen, T., 2004. Multisensory integration: methodological approaches and emerging principles in the human brain. J. Physiol. Paris 98, 191–205.
Calvert, G.A., Campbell, R., Brammer, M.J., 2000. Evidence from functional magnetic resonance imaging of crossmodal binding in the human heteromodal cortex. Curr. Biol. 10, 649–657.
Derrfuss, J., Brass, M., Neumann, J., von Cramon, D.Y., 2005. Involvement of the inferior frontal junction in cognitive control: meta-analyses of switching and Stroop studies. Hum. Brain Mapp. 25, 22–34.
Fingelkurts, A.A., Fingelkurts, A.A., Krause, C.M., Mottonen, R., Sams, M., 2003. Cortical operational synchrony during audio-visual speech integration. Brain Lang. 85, 297–312.
Forman, S.D., Cohen, J.D., Fitzgerald, M., Eddy, W.F., Mintun, M.A., Noll, D.C., 1995. Improved assessment of significant activation in functional magnetic resonance imaging (fMRI): use of a cluster-size threshold. Magn. Reson. Med. 33, 636–647.
Forster, B., Cavina Pratesi, C., Aglioti, S.M., Berlucchi, G., 2002. Redundant target effect and intersensory facilitation from visual-tactile interactions in simple reaction time. Exp. Brain Res. 143, 480–487.
Frassinetti, F., Bolognini, N., Ladavas, E., 2002. Enhancement of visual perception by crossmodal visuo-auditory interaction. Exp. Brain Res. 147, 332–343.
Friston, K.J., Buechel, C., Fink, G.R., Morris, J., Rolls, E., Dolan, R.J., 1997. Psychophysiological and modulatory interactions in neuroimaging. Neuroimage 6, 218–229.
Friston, K.J., Fletcher, P., Josephs, O., Holmes, A., Rugg, M.D., Turner, R., 1998. Event-related fMRI: characterizing differential responses. Neuroimage 7, 30–40.
Friston, K.J., Holmes, A.P., Price, C.J., Buchel, C., Worsley, K.J., 1999. Multisubject fMRI studies and conjunction analyses. Neuroimage 10, 385–396.
Gilbert, C.D., Sigman, M., Crist, R.E., 2001. The neural basis of perceptual learning. Neuron 31, 681–697.
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.
Hertrich, I., Dietrich, S., Ackermann, H., 2011. Cross-modal interactions during perception of audiovisual speech and nonspeech signals: an fMRI study. J. Cogn. Neurosci. 23, 221–237.
Hughes, H.C., Reuter Lorenz, P.A., Nozawa, G., Fendrich, R., 1994. Visual–auditory interactions in sensorimotor processing: saccades versus manual responses. J. Exp. Psychol. Hum. Percept. Perform. 20, 131–153.
Jancke, L., Wustenberg, T., Scheich, H., Heinze, H.J., 2002. Phonetic perception and the temporal cortex. Neuroimage 15, 733–746.
Jones, J.A., Callan, D.E., 2003. Brain activity during audiovisual speech perception: an fMRI study of the McGurk effect. Neuroreport 14, 1129–1133.
Jones, J.A., Jarick, M., 2006. Multisensory integration of speech signals: the relationship between space and time. Exp. Brain Res. 174, 588–594.
Kilian-Hütten, N., Valente, G., Vroomen, J., Formisano, E., 2011. Auditory cortex encodes the perceptual interpretation of ambiguous sound. J. Neurosci. 31, 1715–1720.
McGurk, H., MacDonald, J., 1976. Hearing lips and seeing voices. Nature 264, 746–748.
Meienbrock, A., Naumer, M.J., Doehrmann, O., Singer, W., Muckli, L., 2007. Retinotopic effects during spatial audio-visual integration. Neuropsychologia 45, 531–539.
Mottonen, R., Krause, C.M., Tiippana, K., Sams, M., 2002. Processing of changes in visual speech in the human auditory cortex. Brain Res. Cogn. Brain Res. 13, 417–425.
Myers, E.B., Blumstein, S.E., Walsh, E., Eliassen, J., 2009. Inferior frontal regions underlie the perception of phonetic category invariance. Psychol. Sci. 20, 895–903.
Naumer, M.J., van den Bosch, J.J., 2009. Touching sounds: thalamocortical plasticity and the neural basis of multisensory integration. J. Neurophysiol. 102, 7–8.
Naumer, M.J., Doehrmann, O., Muller, N.G., Muckli, L., Kaiser, J., Hein, G., 2009. Cortical plasticity of audio-visual object representations. Cereb. Cortex 19, 1641–1653.
Nishitani, N., Hari, R., 2002. Viewing lip forms: cortical dynamics. Neuron 36, 1211–1220.
Ojanen, V., Mottonen, R., Pekkola, J., Jaaskelainen, I.P., Joensuu, R., Autti, T., Sams, M., 2005. Processing of audiovisual speech in Broca's area. Neuroimage 25, 333–338.
Olson, I.R., Gatenby, J.C., Gore, J.C., 2002. A comparison of bound and unbound audio-visual information processing in the human cerebral cortex. Brain Res. Cogn. Brain Res. 14, 129–138.
Raizada, R.D., Poldrack, R.A., 2007. Selective amplification of stimulus differences during categorical processing of speech. Neuron 56, 726–740.
Saito, D.N., Yoshimura, K., Kochiyama, T., Okada, T., Honda, M., Sadato, N., 2005. Cross-modal binding and activated attentional networks during audio-visual speech integration: a functional MRI study. Cereb. Cortex 15, 1750–1760.
Schroger, E., Widmann, A., 1998. Speeded responses to audiovisual signal changes result from bimodal integration. Psychophysiology 35, 755–759.
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, A.B., Lau, R., 1997. Sound alters visual motion perception. Nature 385, 308.
Shimojo, S., Shams, L., 2001. Sensory modalities are not separate modalities: plasticity and interactions. Curr. Opin. Neurobiol. 11, 505–509.
Stevenson, R.A., James, T.W., 2009. Audiovisual integration in human superior temporal sulcus: inverse effectiveness and the neural processing of speech and object recognition. Neuroimage 44, 1210–1223.
van Atteveldt, N., Formisano, E., Goebel, R., Blomert, L., 2004. Integration of letters and speech sounds in the human brain. Neuron 43, 271–282.
van Atteveldt, N.M., Formisano, E., Blomert, L., Goebel, R., 2007. The effect of temporal asynchrony on the multisensory integration of letters and speech sounds. Cereb. Cortex 17, 962–974.
Vroomen, J., van Linden, S., de Gelder, B., Bertelson, P., 2007. Visual recalibration and selective adaptation in auditory–visual speech perception: contrasting build-up courses. Neuropsychologia 45, 572–577.
Werner, S., Noppeney, U., 2009. Superadditive responses in superior temporal sulcus predict audiovisual benefits in object categorization. Cereb. Cortex.
Wiersinga Post, E., Tomaskovic, S., Slabu, L., Renken, R., de Smit, F., Duifhuis, H., 2010. Decreased BOLD responses in audiovisual processing. Neuroreport 21, 1146–1151.
Wright, T.M., Pelphrey, K.A., Allison, T., McKeown, M.J., McCarthy, G., 2003. Polysensory interactions along lateral temporal regions evoked by audiovisual speech. Cereb. Cortex 13, 1034–1043.