Examining the McGurk illusion using high-field 7 Tesla functional MRI

Gregor R. Szycik[1], Jörg Stadler[2], Claus Tempelmann[3] and Thomas F. Münte[4]*

[1] Department of Psychiatry, Medical School Hannover, Hannover, Germany
[2] Leibniz Institute for Neurobiology, Magdeburg, Germany
[3] Department of Neurology, University of Magdeburg, Magdeburg, Germany
[4] Department of Neurology, University of Lübeck, Lübeck, Germany

In natural communication speech perception is profoundly influenced by observable mouth movements. The additional visual information can greatly facilitate intelligibility but incongruent visual information may also lead to novel percepts that neither match the auditory nor the visual information as evidenced by the McGurk effect. Recent models of audiovisual (AV) speech perception accentuate the role of speech motor areas and the integrative brain sites in the vicinity of the superior temporal sulcus (STS) for speech perception. In this event-related 7 Tesla fMRI study we used three naturally spoken syllable pairs with matching AV information and one syllable pair designed to elicit the McGurk illusion. The data analysis focused on brain sites involved in processing and fusing of AV speech and engaged in the analysis of auditory and visual differences within AV presented speech. Successful fusion of AV speech is related to activity within the STS of both hemispheres. Our data supports and extends the audio-visual-motor model of speech perception by dissociating areas involved in perceptual fusion from areas more generally related to the processing of AV incongruence.

Keywords: audio-visual integration, McGurk illusion, 7 Tesla, functional magnetic resonance imaging

INTRODUCTION

Audiovisual (AV) integration in the perception of speech is the rule rather than the exception. Indeed, speech perception is profoundly influenced by observable mouth movements. The presence of additional visual (V) information leads to a considerable improvement of intelligibility of auditory (A) input under noisy conditions (Sumby and Pollack, 1954; Schwartz et al., 2004; Ross et al., 2007). Furthermore, lip-movements may serve as a further cue in complex auditory environments, such as the proverbial cocktail-party, and may thus enhance selective listening (Driver, 1996). Recently, it has been suggested that even under auditory-only conditions, speaker-specific predictions and constraints are retrieved by the brain by recruiting face-specific processing areas, in order to improve speech recognition (von Kriegstein et al., 2008).

A striking effect, first described by McGurk and MacDonald (1976), is regularly found when mismatching auditory (e.g., /ba/) and visual (e.g., /ga/) syllables are presented: in this case many healthy persons perceive a syllable neither heard nor seen (i.e., /da/). This suggests that AV integration during speech perception not only occurs automatically but also is more than just the visual modality giving the auditory modality a hand in the speech recognition process. This raises the question as to where such integration takes place in the brain.

A number of recent neuroimaging studies have addressed the issue of AV integration in speech perception and consistently found two different brain areas, the inferior frontal gyrus (IFG) and the superior temporal sulcus (STS) (e.g., Cusick, 1997;

Sekiyama et al., 2003; Wright et al., 2003; Barraclough et al., 2005; Szycik et al., 2008a,b; Brefczynski-Lewis et al., 2009; Lee and Noppeney, 2011; Nath et al., 2011; Nath and Beauchamp, 2011, 2012; Stevenson et al., 2011). As the posterior part of the STS area has been shown to receive input from unimodal A and V cortex (Cusick, 1997) and to harbour multisensory neurons (Barraclough et al., 2005; Szycik et al., 2008b), it can be assumed to be a multisensory convergence site. This is corroborated by the fact that activation of the caudal part of the STS is driven by AV presented syllables (Sekiyama et al., 2003), monosyllabic words (Wright et al., 2003), disyllabic words (Szycik et al., 2008a), and speech sequences (Calvert et al., 2000). Whereas the evidence for a role of the STS area in AV speech processing is growing, it has to be pointed out that it has also been implicated in the AV integration of non-verbal material (e.g., Macaluso et al., 2000; Calvert, 2001; Jäncke and Shah, 2004; Noesselt et al., 2007; Driver and Noesselt, 2008) as well as in diverse other cognitive domains such as empathy (Krämer et al., 2010), biological motion perception and face processing (Hein and Knight, 2008).

Besides the STS, Broca's area, situated on the IFG of the left hemisphere, and its right hemisphere homologue have been implicated in AV speech perception. It can be subdivided into the pars triangularis (PT) and the pars opercularis (PO) (Amunts et al., 1999), Whereas Broca's area has traditionally been deemed to be most important for language production, many neuroimaging studies have suggested multiple functions of this area in language processing [reviewed, e.g., in (Bookheimuer, 2002; Hagoort, 2005) and beyond (Koelsch et al., 2002)].

Significantly elaborating on earlier ideas of Liberman (Liberman and Mattingly, 1989, 1985), Skipper and colleagues (Skipper et al., 2005, 2007) recently introduced the so-called audio-visual-motor integration model of speech perception. In their model they emphasize the role of the PO for speech comprehension by recruiting the mirror neuron system and the motor system. AV speech processing is thought to involve the formation of a sensory hypothesis in STS which is further specified in terms of the motor goal of the articulatory movements established in the PO. In addition to the direct mapping of the sensory input, the interaction of both structures via feedback connections is of paramount importance for speech perception according to this (Skipper et al., 2007).

To further investigate the roles of the IFG and STS in AV speech integration we used naturally spoken syllable pairs with matching AV information (i.e., visual and auditory information /ba//ba/, henceforth BABA, /ga//ga/, GAGA, or /da//da/, DADA) and one audiovisually incongruent syllable pair designed to elicit the McGurk illusion (auditory /ba//ba/ dubbed on visual /ga//ga/, henceforth M-DADA). The study design allowed us to control for the effect of the McGurk illusion, for the visual respective auditory differences of the presented syllables and for task effects. Furthermore, we took advantage of the high signal-to-noise ratio of high-field 7 Tesla functional MRI.

The effect of the McGurk illusion was delineated by contrasting the M-DADA with DADA as well as M-DADA stimuli that gave rise to an illusion with those that did not. Brain areas showing up in these contrasts are considered to be involved in AV fusion. Contrasting stimuli that were physically identical except for the auditory stream (M-DADA vs. GAGA) revealed brain areas involved in the processing of auditory differences, whereas the comparison M-DADA vs. BABA revealed regions concerned with the processing of visual differences.

METHODS

All procedures had been cleared by the ethical committee of the University of Magdeburg and conformed to the declaration of Helsinki.

PARTICIPANTS

Thirteen healthy right-handed native speakers of German gave written informed consent to participate for a small monetary compensation. The dataset of one participant was excluded, as he did not classify any of the target stimuli correctly during the measurement. Thus, 12 participants were included in the analysis (five women, age range 21–39 years).

STIMULI AND DESIGN OF THE STUDY

AV stimuli were spoken by a female native speaker of German with linguistic experience and recorded by a digital camera and a microphone. The speaker kept her face still except for the articulatory movements. Three naturally spoken syllable pairs with matching AV information (AV/ga//ga/, AV/ba//ba/, AV/da//da/, henceforth denoted as GAGA, BABA, DADA) and one syllable pair designed to elicit the McGurk illusion by dubbing the audio stream of BABA onto the video of the face vocalising GAGA (A/ba//ba/V/ga//ga/, henceforth denoted as M-DADA) comprised

the stimulus set. The M-DADA stimulus gave rise to the illusory perception of /da//da/ in many cases. The recorded video was cut into segments of 2 s length (250 × 180 pixel resolution) showing the frontal view of the lower part of speaker's face. The eyes of the speaker were not visible to avoid shifting of the participant's gaze away from the mouth. The mouth of the speaker was shut at the beginning of each video-segment. The vocalization and articulatory movements began 200 ms after the onset of the video.

A slow event-related design was used for the study. Stimuli were separated by a resting period of 10 s duration resulting in a stimulus-onset-asynchrony of 12 s. During the resting period the participants had to focus their gaze on a fixation cross shown at the same position as the speaker's mouth. Stimuli were presented in a pseudo-randomized order with 30 repetitions per stimulus class resulting in the presentation of 120 stimuli total and an experimental duration of 24 min.

To check for the McGurk illusion, participants were required to press one button every time they perceived /da//da/ (regardless of whether this perception was elicited by a DADA or an M-DADA stimulus) and another button whenever they heard another syllable. Thus, to the extent that the M-DADA gave rise to the McGurk illusion, subjects were pressing the /da//da/ button for this stimulus.

Presentation software (Neurobehavioral Systems, Inc.) was used for stimulus delivery. Audio-information was presented binaurally in mono-mode via fMRI compatible electrodynamic headphones integrated into earmuffs to reduce residual background scanner noise (Baumgart et al., 1998). The sound level of the stimuli was individually adjusted to achieve good audibility during data acquisition. Visual stimuli were projected via a mirror system by LCD projector onto a diffusing screen inside the magnet bore.

IMAGE ACQUISITION AND ANALYSIS

Magnetic resonance images were acquired on a 7T MAGNETOM Siemens Scanner (Erlangen, Germany) located in Magdeburg, Germany, and equipped with a CP head coil (*In vivo*, Pewaukee, WI, USA). A total of 726 T2*-weighted volumes covering part of the brain including frontal speech areas and the middle and posterior part of the STS were acquired (TR 2000 ms, TE 24 ms, flip angle 80°, FOV 256 × 192 mm, matrix 128 × 96, from 21–26 slices—depending on specific absorption rate (SAR) limitations, slice thickness 2.0 mm, interslice gap 0.2 mm) (see **Figure 1**). Additionally, a 3D high resolution T1-weighted volume (MPRAGE, TR 2200 ms, TI 1100 ms, flip angle 9°, matrix 192 × 2562, 1 mm isovoxel) was obtained. The subject's head was firmly padded with foam cushions during the entire measurement to avoid head movements.

Analysis and visualization of the data were performed using Brain Voyager QX (Brain Innovation BV, Maastricht, Netherlands) software. First, a correction for the temporal offset between the slices acquired in one scan was applied. For this purpose the data was sinc-interpolated. Subsequently, 3D motion correction was performed by realignment of the entire volume set to the first volume by means of trilinear interpolation. Thereafter, linear trends were removed and a high pass filter was applied
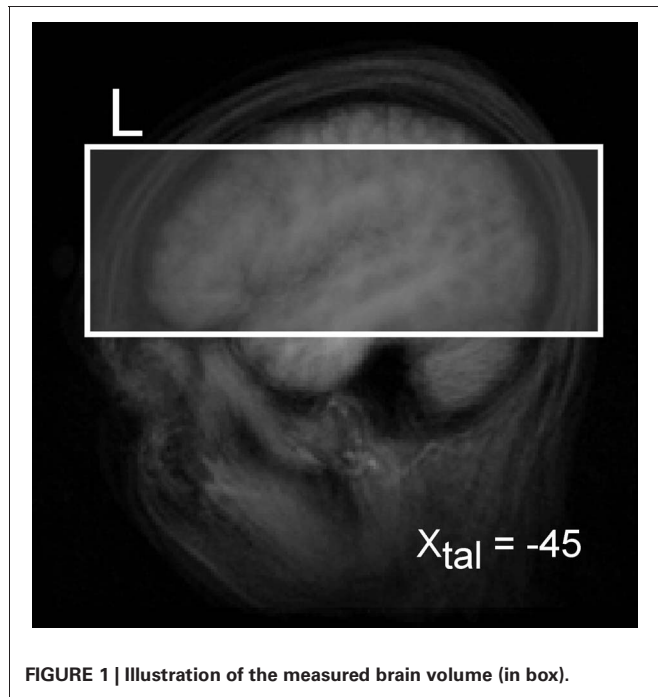
FIGURE 1 | Illustration of the measured brain volume (in box).

Table 1 | Performance data.

Classified as:	Illusion Group ($n = 7$)		No-Illusion Group ($n = 5$)	
	M-DADA	**DADA**	**M-DADA**	**DADA**
/da//da/	43.3 (9.6)	73.3 (13.3)	1.3 (1.4)	96.6 (1.4)
other	53.3 (9.6)	23.3 (15.0)	98.6 (1.4)	3.3 (1.4)
no response	3.3	3.8	–	–

M-DADA and DADA classified as /da-da/ or "other" in percent. Clearly, the illusion worked only in a sub-group of the participants.

resulting in filtering out signals occurring less then three cycles during the whole time course.

Structural and functional data were spatially transformed into the Talairach standard space using a 12-parameter affine transformation. Functional EPI volumes were resampled into 2 mm cubic voxels and then spatially smoothed with an 8 mm full-width half-maximum isotropic Gaussian Kernel to accommodate residual anatomical differences across volunteers.

Prior to the statistical analysis of the fMRI data we divided the participants into two groups according to whether or not they were prone to the McGurk illusion as evidenced by their response pattern. The illusion and no-illusion group comprised seven and five participants, respectively (for details see **Table 1**). The main analyses were conducted for the illusion group.

For the statistical model a fitted design matrix including all conditions of interest was specified using a hemodynamic response function. This function was created by convolving the rectangle function with the model of Boynton et al. (Boynton et al., 1996) using delta = 2.5, tau = 1.25 and $n = 3$. Thereafter a multi-subject fixed effects general linear model (GLM) analysis was used for identification of significant differences in

hemodynamic responses. As regressors of no interest we used overall six translation and rotation vectors derived for each dataset during the 3D motion correction. The statistical analysis included the AV native syllable pairs GAGA, BABA, DADA as well as the M-DADA events. The latter were separated into those events that gave rise to an illusion and those that gave not (see **Table 1**).

Several different linear contrasts were calculated to pinpoint brain regions involved in various aspects of syllable recognition and AV integration:

(a) The effect of the McGurk illusion was delineated by calculating the conjunction of M-DADA, DADA, and M-DADA vs. DADA. Only events for which the participants indicated the perception of /da//da/ were included.
(b) By calculating the conjunction of M-DADA vs. GAGA with M-DADA and GAGA we investigated brain regions involved in the processing of auditory differences within AV stimuli.
(c) In a similar fashion we calculated the conjunction of M-DADA vs. BABA with M-DADA and BABA to identify regions concerned with the processing of visual differences within AV stimuli.
(d) Finally, brain regions involved in AV integration were isolated by contrasting M-DADA stimuli that gave or gave not rise to the McGurk illusion (i.e., the percept /da-da/).

In addition we checked the processing differences between pairs of non-McGurk syllables GAGA, BABA, DADA and the effects of the task by contrasting GAGA, and BABA vs. DADA.

The threshold of $p < 0.005$ was chosen for identification of the activated voxels. Furthermore, only activations involving contiguous clusters of at least 20 functional voxels (corresponding to a volume of 160 mm³) are reported. The centres of mass of suprathreshold regions were localized using Talairach coordinates and the Talairach Daemon tool (Lancaster et al., 2000).

Additionally, we calculated group differences contrasting all M-DADA events (regardless of the response) of the illusion group (i.e., participants who had the McGurk illusion) against the no-illusion group. For this analysis the threshold was set to $p < 0.005$ (FDR-corrected) and minimal cluster volume was set to 20 voxels.

RESULTS

The performance indicated that participants attentively processed the stimuli and that they experienced the McGurk illusion in a high proportion of cases (see **Table 1**).

The contrast M-DADA vs. DADA (see **Figure 2A**, red activation) revealed clusters in and near the STS of the right hemisphere. Additionally, left and right insula showed greater activity for the M-DADA condition in comparison to the DADA condition. There were no significantly activated areas showing stronger activity for the DADA condition in comparison to the M-DADA condition.

The comparison of brain activity for the M-DADA classified as /da-da/ vs. classified as non-/da-da/ **Figure 2A** (yellow/orange activations) revealed areas involved in effective AV integration. Two clusters in the STG region were found. We did not find any significant activity that was stronger for M-DA events not
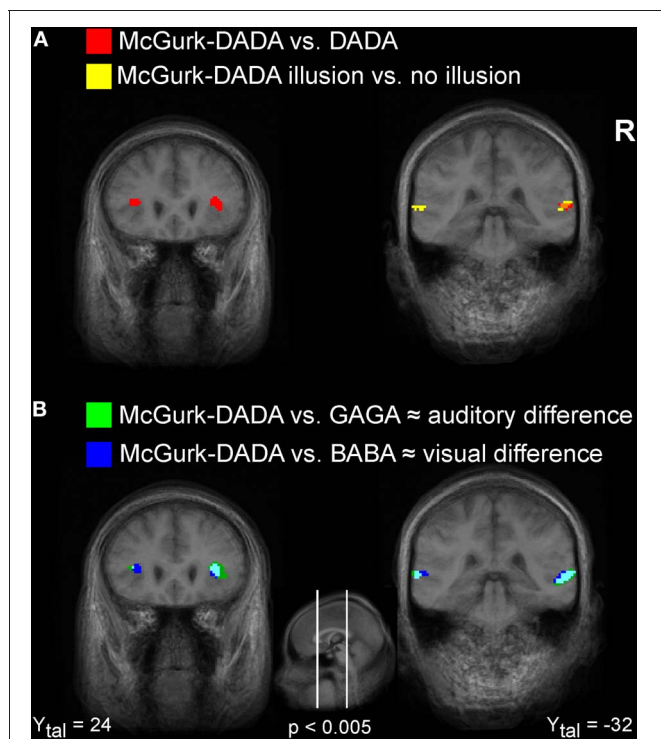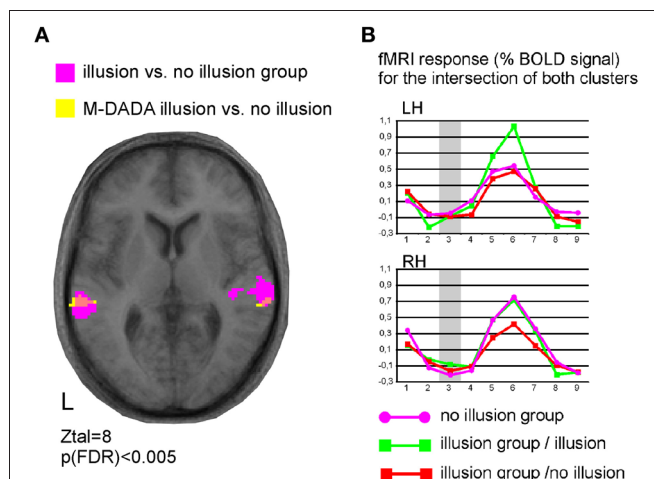
FIGURE 2 | Visualization of fMRI activations. Illustration of activated clusters rendered on a T1weighted average brain image. **(A)** Areas showing stronger activity for M-DADA in comparison to DADA are colored in red. Areas showing stronger activity for M-DADA with illusory perception relative to M-DADA without illusion are colored in yellow. Overlapping areas are shown in orange. **(B)** Areas showing stronger activity for the M-DADA with illusion in comparison to naturally spoken GAGA, reflecting auditory differences, are shown in green. Areas showing stronger activity for the M-DADA with illusion in comparison to naturally spoken BABA, reflecting visual differences, are shown in blue. L = left, R = right, Xtal = Talairach coordinate *x*, Ytal = Talairach coordinate *y*, Ztal = Talairach coordinate *z*.

FIGURE 3 | Functional specialization within the superior temporal cortex. (A) Illustration of activated clusters, L = left, Ztal = Talairach coordinate *z*, pink colour indicates the overlap of both contrasts. **(B)** Time courses for the M-DADA events extracted from both overlap clusters for the group of participants with McGurk illusion (*n* = 7) and the group with no-illusion (*n* = 5). RH = right and LH = left hemisphere.

classified as /da-da/ in comparison to events with occurred integration.

Figure 2B illustrates brain regions involved in the processing of auditory (green) and visual (blue) differences during AV stimulation. The comparison of M-DADA vs. GAGA events (auditory difference) yielded activation in the insula, STG of both hemispheres and additionally in the cuneus of the right hemisphere (**Table 2**). The activated areas of STG occupy the same spatial location as the STG clusters identified by contrasting M-DADA against DA and M-DADA classified as /da-da/ vs. non-/da-da/ described above. There were no significant activations for contrasting GAGA against M-DADA events.

The areas processing visual differences were identified by contrasting M-DADA vs. BABA events. There were two clusters in the STG respectively insula and cuneus of both hemispheres and one cluster in the right superior and IFG and in the middle occipital gyrus (**Table 2**). Thus, the STS region is involved in matching visual and auditory information.

Brain responses to the M-DADA stimuli differed between the Illusion and the non-Illusion group (**Figure 3**, **Table 2** and **3**).

As the illusion group showed greater activation of the STS, this again underscores the importance of this area for the integration of divergent AV information into a coherent percept. Moreover, within the illusion group those M-DADA events leading to an illusion showed a greater response in the STS than M-DADA events that did not. Further analyses revealed no significant activations when non-McGurk stimuli were subjected to pair-wise comparisons.

DISCUSSION

The present study took advantage of very high-field functional MRI to define the brain areas involved in the McGurk illusion during the AV integration of speech. Importantly, different percepts revealed by the same stimuli evoked different brain activity in the STS region corroborating its role in the AV processing of speech.

This activation pattern is very similar to that found in other studies comparing AV congruent and incongruent speech stimuli (Ojanen et al., 2005; Pekkola et al., 2005, 2006; Szycik et al., 2008a; Brefczynski-Lewis et al., 2009; Lee and Noppeney, 2011; Nath et al., 2011; Stevenson et al., 2011). These previous studies employed AV speech stimuli with either congruent [e.g., auditory: hotel, visual: hotel, (Szycik et al., 2008a)] or incongruent visual information (e.g., auditory: hotel, visual: island) but were not designed to elicit a McGurk-illusion. Therefore, the activity in the fronto-temporo-parietal network observed there can not be entirely due to the AV integration processes that give rise to the illusion. Rather, the increased activity for the AV speech percept appears to mainly reflect increased neural processing caused by the divergent input information provided by incongruent AV speech. Miller and D'Esposito (2005) contrasted synchronous and asynchronous AV speech (identical vowel-consonant-vowel syllables in the auditory and visual domain with different temporal

Table 2 | Tabulated results of the fMRI analysis (only participants with illusion) Talairach coordinates for the center of mass of the particular cluster and its size in 3 mm.

Regions	≈BA	Hemisphere	Talairach center of mass			Cluster size (mm^3)
			x	y	z	
M-DADA > DADA						
Cingulate Gyrus	23	L	−1	−22	29	168
Precuneus	31	R	4	−67	27	712
Insula	13	L	−32	23	13	304
Insula	13	R	34	24	12	632
Superior Temporal Gyrus	22	R	55	−28	7	1016
Cuneus	17	R	17	−95	3	648
M-DADA (PERCEPT /DA-DA/ VS. NON-/DA-DA/)						
Superior Temporal Gyrus	22	L	−58	−28	7	488
Superior Temporal Gyrus	42	R	60	−31	10	384
M-DADA > BABA ≈ VISUAL DIFFERENCE						
Superior Temporal Gyrus	7	L/R	0	−66	32	1496
Superior Frontal Gyrus	10	R	33	56	24	504
Inferior Frontal Gyrus	9	R	43	15	26	184
Insula	13	R	33	24	11	384
Insula	13	L	−31	24	11	312
Superior Temporal Gyrus	42	R	58	−32	6	1060
Superior Temporal Gyrus	42	L	−60	−31	7	376
Middle Occipital Gyrus	18	R	22	−98	8	656
M-DADA > GAGA ≈ AUDITORY DIFFERENCE						
Cuneus	7	R	3	−66	32	3240
Insula	13	R	35	24	11	1096
Insula	13	L	−35	21	13	248
Superior Temporal Gyrus	22	R	59	−33	6	1012
Superior Temporal Gyrus	22	L	−63	−32	7	176

BA = approximate Brodmann's area; L = Left hemisphere, R = Right hemisphere.

Table 3 | Comparison Illusion vs. No-Illusion group, M-DA stimuli All M-DADA stimuli were entered into the analysis regardless of whether they elicited an illusion or not.

Regions	≈BA	Hemisphere	Talairach center of mass			Cluster size (mm^3)
			x	y	z	
Superior Temporal Gyrus	22	L	−55	−38	13	1960
Superior Temporal Gyrus	42	R	62	−28	13	1828

Talairach coordinates for the center of mass of the particular cluster and its size in 3 mm (qFDR < 0.05). BA = approximate Brodmann's area; L = Left hemisphere, R = Right hemisphere.

offsets) and searched for brain areas involved in the perceptual fusion of the information from the two modalities as evidenced by the participants' report of perceiving both inputs as synchronous. The left STS and left Heschl's gyrus showed increased activity with perceptual fusion defined as perception of synchronous AV streams.

By comparing the M-DADA events that led to an illusion (i.e., were perceived as /da-da/) vs. M-DADA events that did not, brain areas closely related to the fusion processes underlying the illusion could be delineated. These areas were found bilaterally in the posterior part of STS. Importantly, in this contrast physically identical stimuli were compared. Also, trials giving rise to the illusion and trials that did not occurred about equally often (c.f., **Table 1**) making this a valid comparison. This suggests that the bilateral STS is involved in the perceptual binding processes leading to the McGurk perceptual fusion effect. This region corresponds well to previous results that have suggested an involvement of STS in the integration of auditory as well as visual speech stimuli (Beauchamp, 2005; Bernstein et al., 2008; Szycik et al., 2008b). Comparing the activity to M-DADA stimuli

for subjects that were prone to the illusion and subjects that were not susceptible to the illusion again pointed to the STS region bilaterally as a major site for AV integration. This result is similar to one obtained very recently by Nath and Beauchamp (2012). In their study only the left STS showed a significant effect of both susceptibility to the illusion and stimulus condition. As in the present investigation, the amplitude of the response in the left STS was greater for subjects that were likely to perceive the McGurk effect. Nath and Beauchamp concluded that the left STS is a key locus for interindividual differences in speech perception, a conclusion also supported by the present data.

Comparing M-DADA (illusion) events vs. BABA (GAGA) events enabled us to identify brain regions involved in processing of visual (auditory) differences in AV presented speech (**Figure 2B**). As the contrasts between the plain DADA stimuli with either BABA or GAGA, as well as that between GAGA and DADA did not reveal any significant activations, we conclude that the M-DADA vs. BABA (GAGA) comparison reflects brain activity-related to the integration of visual and auditory differences in AV presented speech. Auditory as well as visual differences were processed bilaterally in the anterior insula and the STS. Visual differences were additionally processed in right frontal brain structures indicating these impact for mapping of the visual input onto the auditory. The impact of right frontal brain structures on AV speech processing has previous been shown, for example in participants suffering from schizophrenia (Szycik et al., 2009).

Our results fit with the interpretation of the functions of the STS presented by Hein and Knight (2008). Taking into account neuroimaging studies focusing on theory of mind, AV integration, motion processing, speech processing, and face processing, these authors suggest that the anterior STS region is mainly involved in speech processing, whereas the posterior portion is recruited by cognitive demands imposed by the different task domains. They further argue against distinct functional subregions in the STS and adjacent cortical areas and rather suggest that the same STS region can serve different cognitive functions as a flexible component in networks with other brain regions.

Keeping this in mind, our pattern of results is in agreement with a recent model emphasizing the role of the STS region for AV speech integration. In the audio-visual-motor model of speech perception of Skipper and coworkers (Skipper et al., 2007), the observation of mouth movements leads to the formation of a "hypothesis" in multisensory posterior superior temporal areas which is specified in terms of the motor goal of that movement in the PO of the inferior frontal cortex which in turn maps onto more explicit motor commands in the premotor cortex. From these motor commands predictions are generated about the auditory (fed back from the premotor cortex to the posterior STS region) and somatosensory (fed back from the premotor cortex to the somatosensory cortex, supramarginal gyrus, and STS region) consequences such commands would have. These feedback influences shape the speech perception in the STS region by supporting a particular interpretation of the input. Importantly, both, the auditory input and the visual input, are used for the
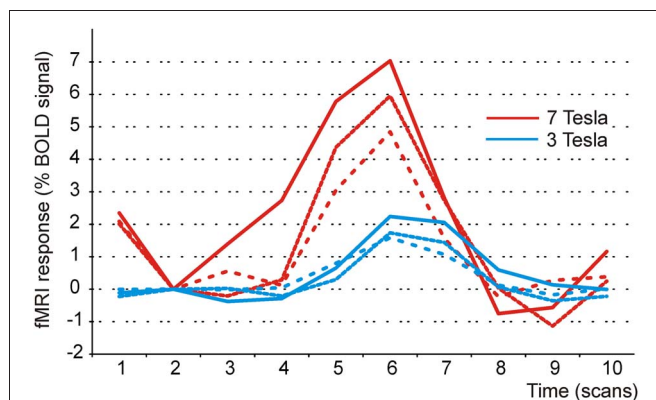
FIGURE 4 | BOLD response in the auditory cortex to audiovisual speech stimuli at 3 and 7 Tesla. Illustration of signal changes in the auditory cortex elicited by audiovisual stimuli at 7 T and 3 T. Shown are three participants of the current study and three randomly chosen participants taken from Szycik et al. (2008a). In this study audiovisually presented bisyllabic nouns were used. Clearly, signal changes are considerably more pronounced at 7 T.

generation of hypotheses. For McGurk stimuli, the hypotheses generated from the auditory and visual input are in conflict and are integrated (fused) in the posterior STS region. Our results are thus in line with the Skipper et al. model. They further suggest a crucial role of the STS for perception of ambiguous AV speech.

On a more general level, the current data also speak to the utility of 7 T imaging for cognitive neuroscience. To get an idea about the magnitude of the effects obtained in the current slow event-related design at 7 T, we computed the BOLD signal change in percent in a ROI covering the transverse temporal gyrus in three randomly chosen participants (shown in red in **Figure 4**). In addition, we did the same for three randomly chosen participants (shown in blue) from an earlier study (Szycik et al., 2008a) which had used a similar slow-event-related design at 3 Tesla (TR 2000 ms, TE 30 ms, flip angle 80°, FOV 224 mm matrix 642, 30 slices, slice thickness 3.5 mm interslice gap 0.35 mm). Stimuli in the 3 Tesla study comprised short video-clips of 2 s (AV presentation of bisyllabic German nouns, no McGurk illusion intended) followed by a 16 s rest period. As the SOA between the critical stimuli was somewhat longer in the 3 T study (18 s, current study 12 s SOA), the settings in that study were even more favourable for observing large signal changes. In spite of this, the signal change amounted to 5–7 percent in the 7 T and to only 1–2 percent in the 3 T study.

This suggests that moving from 3 to 7 Tesla allows to either increase signal-to-noise ratio or reduce the number of trials needed. The latter option might be particularly interesting for investigations of patients or in psycholinguistics, where often only small numbers of stimuli are available.

REFERENCES

Amunts, K., Schleicher, A., Bürgel, U., Mohlberg, H., Uylings, H. B. M., and Zilles, K. (1999). Broca's region revisited: cytoarchitecture and intersubject variability. *J. Comp. Neurol.* 412, 319–341.

Barraclough, N. E., Xiao, D., Baker, C. I., Oram, M. W., and Perrett, D. I. (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., and Scheich, H. (1998). Electrodynamic headphones and woofers for application in magnetic resonance imaging scanners. *Med. Phys.* 25, 2068–2070.

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

Bernstein, L. E., Auer, E. T., Wagner, M., and Ponton, C. W. (2008). Spatiotemporal dynamics of audiovisual speech processing. *Neuroimage* 39, 423–435.

Bookheimuer, S. Y. (2002). Functional MRI of Language: new approaches to understanding the cortical organization of semantic processing. *Annu. Rev. Neurosci.* 25, 151–188.

Boynton, G. M., Engel, S. A., Glover, G. H., and Heeger, D. J. (1996). Linear systems analysis of functional magnetic resonance imaging in human V1. *J. Neurosci.* 16, 4207–4221.

Brefczynski-Lewis, J., Lowitszch, S., Parsons, M., Lemieux, S., and Puce, A. (2009). Audiovisual non-verbal dynamic faces elicit converging fMRI and ERP responses. *Brain Topogr.* 21, 193–206.

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., and Brammer, M. J. (2000). Evidence from functional magnetic resonance imaging of crossmodal binding in the human heteromodal cortex. *Curr. Biol.* 10, 649–657.

Cusick, C. G. (1997). "The superior temporal polysensory region in monkeys," in *Cerebral Cortex: Extrastriate Cortex in Primates*, eds K. S. Rockland, J. H. Kaas, and A. Peters (New York, NY: Springer), 435–468.

Driver, J. (1996). Enhancement of selective listening by illusory mislocation of speech sounds due to lip-reading. *Nature* 381, 66–68.

Driver, J., and Noesselt, T. (2008). Multisensory interplay reveals crossmodal influences on "sensory-specific" brain regions, neural responses, and judgments. *Neuron* 57, 11–23.

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

Hein, G., and Knight, R. T. (2008). Superior temporal sulcus–It's my area: or is it? *J. Cogn. Neurosci.* 20, 2125–2136.

Jäncke, L., and Shah, N. J. (2004). Hearing syllables by seeing visual stimuli. *Eur. J. Neurosci.* 19, 2603–2608.

Koelsch, S., Gunter, T. C., Cramon, D. Y., Zysset, S., Lohmann, G., and Friederici, A. D. (2002). Bach speaks: a cortical "language-network" serves the processing of music. *Neuroimage* 17, 956–966.

Krämer, U. M., Mohammadi, B., Doñamayor, N., Samii, A., and Münte, T. F. (2010). Emotional and cognitive aspects of empathy and their relation to social cognition–an fMRI-study. *Brain Res.* 1311, 110–120.

Lancaster, J. L., Woldorff, M. G., Parsons, L. M., Liotti, M., Freitas, C. S., Rainey, L., Kochunov, P. V., Nickerson, D., Mikiten, S. A., and Fox, P. T. (2000). Automated Talairach atlas labels for functional brain mapping. *Hum. Brain Mapp.* 10, 120–131.

Lee, H., and Noppeney, U. (2011). Physical and perceptual factors shape the neural mechanisms that integrate audiovisual signals in speech comprehension. *J. Neurosci.* 31, 11338–11350.

Liberman, A. M., and Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition* 21, 1–36.

Liberman, A. M., and Mattingly, I. G. (1989). A specialization for speech perception. *Science* 243, 489–494.

Macaluso, E., Frith, C. D., and Driver, J. (2000). Modulation of human visual cortex by crossmodal spatial attention. *Science* 289, 1206–1208.

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

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

Nath, A. R., and Beauchamp, M. S. (2011). Dynamic changes in superior temporal sulcus connectivity during perception of noisy audiovisual speech. *J. Neurosci.* 31, 1704–1714.

Nath, A. R., and Beauchamp, M. S. (2012). A neural basis for interindividual differences in the McGurk effect, a multisensory speech illusion. *Neuroimage* 59, 781–787.

Nath, A. R., Fava, E. E., and Beauchamp, M. S. (2011). Neural correlates of interindividual differences in children's audiovisual speech perception. *J. Neurosci.* 31, 13963–13971.

Noesselt, T., Rieger, J. W., Schoenfeld, M. A., Kanowski, M., Hinrichs, H., Heinze, H. J., and Driver, J. (2007). Audiovisual temporal correspondence modulates human multisensory superior temporal sulcus plus primary sensory cortices. *J. Neurosci.* 27, 11431–11441.

Ojanen, V., Möttönen, R., Pekkola, J., Jääskeläinen, I. P., Joensuu, R., Autti, T., and Sams, M. (2005). Processing of audiovisual speech in Broca's area. *Neuroimage* 25, 333–338.

Pekkola, J., Laasonen, M., Ojanen, V., Autti, T., Jääskeläinen, I. P., Kujala, T., and Sams, M. (2005). 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., Jääskeläinen, I. P., Möttönen, R., and Sams, M. (2006). Attention to visual speech gestures enhances hemodynamic activity in the left planum temporale. *Hum. Brain Mapp.* 27, 471–477.

Ross, L. A., Saint-Amour, D., Leavitt, V. M., Javitt, D. C., and Foxe, J. J. (2007). Do you see what I am saying? Exploring visual enhancement of speech comprehension in noisy environments. *Cereb. Cortex* 17, 1147–1153.

Schwartz, J. L., Berthommier, F., and Savariaux, C. (2004). Seeing to hear better: evidence for early audio-visual interactions in speech identification. *Cognition* 93, B69–B78.

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

Skipper, J. I., Nusbaum, H. C., and Small, S. L. (2005). Listening to talking faces: motor cortical activation during speech perception. *Neuroimage* 25, 76–89.

Skipper, J. I., van Wassenhove, V., Nusbaum, H. C., and Small, S. L. (2007). Hearing lips and seeing voices: how cortical areas supporting speech production mediate audiovisual speech perception. *Cereb. Cortex* 17, 2387–2399.

Stevenson, R. A., van Der Klok, R. M., Pisoni, D. B., and James, T. W. (2011). Discrete neural substrates underlie complementary audiovisual speech integration processes. *Neuroimage* 55, 1339–1345.

Sumby, W. H., and Pollack, I. (1954). Visual contribution to speech intelligibility in noise. *J. Acoust. Soc. Am.* 26, 212–215.

Szycik, G. R., Jansma, H., and Münte, T. F. (2008a). Audiovisual integration during speech comprehension: an fMRI study comparing ROI-based and whole brain analyses. *Hum. Brain Mapp.* 30, 1990–1999.

Szycik, G. R., Tausche, P., and Münte, T. F. (2008b). A novel approach to study audiovisual integration in speech perception: localizer fMRI and sparse sampling. *Brain Res.* 1220, 142–149.

Szycik, G. R., Münte, T. F., Dillo, W., Mohammadi, B., Samii, A., Emrich, H. M., and Dietrich, D. E. (2009). Audiovisual integration of speech is disturbed in schizophrenia: an fMRI study. *Schizophr. Res.* 110, 111–118.

von Kriegstein, K., Dogan, O., Grüter, M., Giraud, A. L., Kell, C. A., Grüter, T., Kleinschmidt, A., and Kiebel, S. J. (2008). Simulation of talking faces in the human brain improves auditory speech recognition. *Proc. Nat. Acad. Sci. U.S.A.* 105, 6747–6752.

Wright, T. M., Pelphrey, K. A., Allison, T., McKeown, M. J., and McCarthy, G. (2003). Polysensory interactions along lateral temporal regions evoked by audiovisual speech. *Cereb. Cortex* 13, 1034–1043.