The following full text is a publisher's version.

For additional information about this publication click this link.
http://hdl.handle.net/2066/149936

Please be advised that this information was generated on 2017-12-20 and may be subject to change.
Preferenee for Audiovisual Speech Congrueney in Superior Temporal Cortex

Claudia S. Lüttke, Matthias Ekman, Marcel A. J. van Gerven, and Floris P. de Lange

Abstract

Auditory speech perception can be altered by concurrent visual information. The superior temporal cortex is an important combining site for this integration process. This area was previously found to be sensitive to audiovisual congruency. However, the direction of this congruency effect (i.e., stronger or weaker activity for congruent compared to incongruent stimulation) has been more equivocal. Here, we used fMRI to look at the neural responses of human participants during the McGurk illusion—in which auditory /aba/ and visual /aga/ inputs are fused to perceived /ada/—in a large homogenous sample of participants who consistently experienced this illusion. This enabled us to compare the neuronal responses during congruent audiovisual stimulation with incongruent audiovisual stimulation leading to the McGurk illusion while avoiding the possible confounding factor of sensory surprise that can occur when McGurk stimuli are only occasionally perceived. We found larger activity for congruent audiovisual stimuli than for incongruent (McGurk) stimuli in bilateral superior temporal cortex, extending into the primary auditory cortex. This finding suggests that superior temporal cortex prefers when auditory and visual input support the same representation.

INTRODUCTION

Speech perception relies not only on the incoming auditory signal but also incorporates visual information, such as mouth movements (Grant & Seitz, 2000). Observing the mouth of a speaker helps in interpreting the auditory signal in a conversation, especially in noisy conditions (Ross, Saint-Amour, Leavitt, Javitt, & Foxe, 2007). Visual information can even alter the interpretation of auditory input, as illustrated by the McGurk illusion, where the combination of a visual /ga/ and auditory /ba/ is perceived as /da/ (McGurk & MacDonald, 1976).

It has been shown that the superior temporal cortex (STC) plays an important role as a combining site for audiovisual information (Beauchamp, Nath, & Pasalar, 2010; Calvert, Campbell, & Brammer, 2000). However, although several studies have documented a modulation of activity in the STC by audiovisual congruency, the direction of this interaction is less clear. Several studies find STC to be more active for congruent than incongruent audiovisual stimulation (Van Atteveldt, Formisano, Goebel, & Blomert, 2004; Calvert et al., 2000). However, others have documented the opposite pattern of results, that is, more activity for incongruent audiovisual input. This has been found both when the incongruent input is not merged into a unitary percept (Baum, Martin, Hamilton, & Beauchamp, 2012; Nath & Beauchamp, 2012; Noppeney, Josephs, Hocking, Price, & Friston, 2008) and when inputs are merged, as during the McGurk illusion (Nath & Beauchamp, 2012; Szycik, Stadler, Tempelmann, & Münte, 2012).

What could be an explanation for this discrepancy? One of the contributing factors could be differences in salience between incongruent and congruent audiovisual input (Baum et al., 2012). The infrequent experience of conflicting input from different senses potentially attracts participants’ attention (Baldi & Itti, 2010), thereby increasing neural activity in the involved sensory areas (Den Ouden, Friston, Daw, McIntosh, & Stephan, 2009; Loose, Kaufmann, Auer, & Lange, 2003; Jaencke, Mirzazade, & Shah, 1999). In a similar vein, the surprise associated with incongruent McGurk trials that are not integrated into a coherent percept may conflate multisensory integration with sensory surprise, which is also known to boost neural activity in the relevant sensory areas during multisensory integration (Lee & Noppeney, 2014). In several studies reporting stronger STC activity for McGurk stimuli, participants perceived the illusion on less than half of the trials (Nath & Beauchamp, 2012; Szycik et al., 2012), rendering them more surprising than congruent stimuli. Indeed, there is large interindividual variability in the propensity to merge incongruent audiovisual input into a coherent percept during McGurk trials (Nath & Beauchamp, 2012). By focusing on individuals who are prone to the McGurk illusion, one can control for sensory surprise because they are less aware of the incongruent stimulation than individuals who only perceive the illusion occasionally.

The aim of this study was to characterize neural activity differences between congruent and merged, incongruent audiovisual input while avoiding the possible confound
of sensory surprise. To this end, we studied a homoge-
nous sample of preselected participants who consistently 
perceive the McGurk illusion. To preview, we observed a 
stronger response in the STC for congruent audiovisual 
stimulation compared to incongruent McGurk trials, in 
line with the hypothesis that the STC is more strongly 
driven by concurrent audiovisual stimulation by the same 
auditory and visual content (Van Atteveldt, Blau, Blomert, 
& Goebel, 2010).

METHODS

Participants

Before the neuroimaging experiment, participants were 
screened for their propensity to perceive the McGurk 
ilusion. In total, 55 right-handed healthy volunteers 
(44 women, age range = 18–30 years) participated in 
the behavioral screening. We selected participants who 
perceived the McGurk illusion on at least four of the 
six McGurk videos (i.e., reported /ada/ or /ata/ for a stim-
ulus in which the auditory signal was /aba/ and the visual 
signal was /aga/). The 27 participants who met this cri-
terion took part in the fMRI study. Four of the selected 
participants were excluded from the analysis because of 
an insufficient number of McGurk illusions in the scanner 
( < 75%). The remaining 23 participants were included in 
the analysis (20 women, age range = 19–30 years). All 
participants had normal or corrected-to-normal vision 
and gave written informed consent in accordance with 
the institutional guidelines of the local ethical committee 
(CMO region Arnhem-Nijmegen, the Netherlands) and 
were either financially compensated or received study 
credits for their participation.

Stimuli

The audiovisual stimuli showed the lower part of the face 
of a speaker uttering syllables (see Figure 1). To this end, 
a female speaker was recorded with a digital video cam-
era in a soundproof room while uttering /aba/, /ada/, and 
/aga/. The videos were edited in Adobe Premier Pro CS6 
such that the mouth was always located in the center of 
the screen to avoid eye movements between trials. After 
editing, each video started and ended with a neutral 
mouth slightly opened such that participants could not 
distinguish the videos based on the beginning of the 
video but only by watching the whole video. The stimuli 
were presented using the Presentation software (www. 
neurobs.com). All videos were 1000-msec long with a 
total sound duration of 720 msec. Only the lower part 
of the face from nose to chin was visible in the videos 
to prevent participants’ attention being drawn away from 
the mouth to the eyes. The McGurk stimuli were created 
by overlaying /aga/ movies to the sound of an /aba/ video. 
In total, there were 18 videos, three for every condition 
(audiovisual /aba/, audiovisual /aga/, McGurk, auditory /aba/
, auditory /ada/, auditory /aga/). During the audiovisual trials, 
McGurk stimuli (auditory /aba/ overlaid on /aga/ video) or 
congruent /aba/ or /aga/ stimuli were shown. We also 
included “auditory-only” trials, during which only a static 
image of the face (first frame of the video showing a slightly 
opened mouth) was presented whereas /aba/, /ada/, or /aga/ 
was presented to the participants via MR compatible in-ear 
headphones. A comfortable, but sufficiently loud, volume 
was calibrated for each participant before the start of the 
experiment. Visual stimuli were presented on a black 
background using a projector (60 Hz refresh rate, 1024 × 
768 resolution) located at the rear of the scanner bore 
and viewed through a mirror yielding 6° of horizontal 
and 7° of vertical visual angle.

Procedure

On each trial, audiovisual and auditory stimuli were pre-
sented for 1 sec. Participants had 4.5–6.5 sec after each 
stimulus before a new stimulus appeared on the screen to 
report in a three alternative forced-choice fashion what 
they had heard. They were, however, instructed to always 
focus and attend to the mouth. They had to respond as 
fast and as accurately as possible with their right index 
(/aba/), middle (/ada/), and ring finger (/aga/) using an 
MRI-compatible button box. In-between stimuli, partici-
pants fixated with their eyes on a gray fixation cross in 
the center of the screen where the mouth appears during 
stimuli presentation to minimize eye movements. All stim-
uli were randomly presented in an event-related design 
distributed over six blocks. Every stimulus was repeated 
23 times, yielding 414 trials in total. Additionally, 10 null
events per block (blank screen from the ISI) were presented for 6–8 sec each throughout the experiment. Before the experiment, participants practiced the task in the scanner (six practice trials, one for each condition). In total, the fMRI experiment lasted approximately 2 hr. At the end of the experiment, we also ran a functional localizer to determine regions that were more responsive to auditory syllables than scrambled versions of the syllables. All results reported here refer to the data acquired during the main task.

**fMRI Data Acquisition**

The functional images were acquired with a 3-T Skyra MRI system (Siemens, Erlangen, Germany) using a continuous T2*-weighted gradient-echo EPI sequence (29 horizontal slices, flip angle = 80°, field of view = 192 × 192 × 59 mm, voxel size = 2 × 2 × 2 mm, repetition time/echo time = 2000/30 msec). The structural image was collected using a T1-weighted MPRAGE sequence (flip angle = 8°, field of view = 192 × 256 × 256 mm, voxel size 1 × 1 × 1, repetition time = 2300 msec).

**fMRI Data Analysis**

Analyses were performed using Statistical Parametric Mapping (www.fil.ion.ucl.ac.uk/spm/software/spm8/, Wellcome Trust Centre for Neuroimaging, London, UK). The first five volumes were discarded to allow for scanner equilibration. During preprocessing, functional images were realigned to the first image, slice time corrected to the onset of the first slice, coregistered to the anatomical image, smoothed with a Gaussian kernel with an FWHM of 6 mm, and finally normalized to a standard T1 template image. A high-pass filter (cutoff = 128 sec) was applied to remove low-frequency signals. The preprocessed fMRI time series were analyzed on a subject-by-subject basis using an event-related approach in the context of a general linear model. We modeled the six conditions (three audiovisual, see Figure 1, and three auditory) separately for the six scanning sessions. Unfused McGurk trials (i.e., McGurk trials in which participants provided an “aba” or “aga” response) were included in the model as a separate regressor of no interest. Six motion regressors related to translation and rotation of the head were included in the model. We assessed the effect of congruency between auditory and visual input by contrasting congruent audiovisual input (/aba/ and /aga/) with incongruent input (visual /aga/ and auditory /aba/ culminating in the McGurk illusion of /ada/). An inclusive masking procedure was used to ensure that all activity differences pertained to areas that were more active during audiovisual stimulation than during baseline at a statistical threshold of p < .001. The false alarm rate was controlled by whole-brain correction of p values at the cluster level (FWE p < .05, based on an auxiliary voxel level threshold of p < .001). To anatomically define the primary auditory cortices, we used the anatomical toolbox implemented in SPM (Eickhoff et al., 2007).

**RESULTS**

**Behavioral Results**

Because of our participant selection criteria (see Methods), participants reported /ada/ percepts almost all the time (94% ± 6%, mean ± SD) following the simultaneous presentation of auditory /aba/ and visual /aga/, suggesting multisensory fusion (see Figure 1). Performance was at ceiling level for the audiovisual congruent stimuli, for both audiovisual /aba/ (97%) and audiovisual /aga/ (99%), which indicates that participants maintained attention to the stimuli over the course of the experiment. A repeated-measures ANOVA showed that the RTs for the fused McGurk stimuli were on average larger (546 ± 204 msec, mean ± SD) than for the congruent audiovisual stimuli (446 ± 193 msec, mean ± SD; F(1, 22) = 20.16, p < .0001), suggesting that perceptual decisions about incongruent stimuli were more difficult. A debriefing questionnaire completed after the fMRI session indicated that participants did not realize that their percepts were affected by perceptual fusion of incongruent auditory and visual signals.

**Neuroimaging Results**

We compared BOLD activity during McGurk stimuli (visual /aga/ and auditory /aba/) with audiovisual congruent stimulation (audiovisual /aba/ and /aga/). We found that both left and right superior temporal gyri were more active when auditory and visual inputs matched than during McGurk illusions (see Table 1 and Figure 2). The anatomical location of activity differences was comparable to STC regions that have been found to be responsive to audiovisual congruency (Van Atteveldt et al., 2010). The left STC cluster extended to cyto-architectonically defined (Rademacher et al., 2001) early auditory cortical areas (19.4% in TE1.0, 23.6% in TE1.1), whereas the right cluster fell dorsal and rostral to early auditory cortex. A power analysis that controls for autocorrelations between voxels (PowerMap; sourceforge.net/projects/powermap/) yielded an average effect size of 0.50 (Cohen’s d) for the two superior temporal clusters (Joyce & Hayasaka, 2012).

Different phonemes were perceived by the participants during congruent audiovisual stimuli (/aba/ or /aga/) and McGurk illusions (/ada/). To rule out that the activity differences in STC were due to the differing identity of the percept between conditions, we carried out a control analysis in an independent data set in which participants listened to these three sounds. We found no difference in activity for aba/aga versus ada in either the left or right
STC activity is particularly strongly activated by congruent multisensory input. In the following we will discuss and interpret our findings in relation to the conflicting results in the literature on multisensory integration.

**Different Directions of Congruency Effects in STC**

Our finding of reduced STC activity for McGurk stimuli is in apparent contradiction to some previous studies where STC was found to be more active for incongruent McGurk stimuli than congruent audiovisual stimulation (Baum et al., 2012; Nath & Beauchamp, 2012; Szycik et al., 2012). However, in these studies, participants did not consistently perceive the McGurk illusion. In other words, only occasionally they merged the incongruent auditory and visual inputs into a unified /ada/ percept. Therefore, the process of merging might be relatively surprising for them and thereby attract attention. The surprising and attention-grabbing nature of this merging process might confound these previous findings, as auditory attention may have resulted in increased activity in auditory cortex and STC (Jaencke et al., 1999). In the current study, participants consistently fused incongruent inputs. Therefore, the process of merging the two senses is less surprising and should attract less attention than in individuals where the merging is more the exception rather than the rule. When avoiding the possible confounding factor of sensory surprise, we observed the opposite pattern, that is, less activity in STC during McGurk stimuli than during congruent stimulation.

The involvement of different subregions within STC that favor congruent or incongruent stimulation might be another possible explanation for the conflicting findings. A study on the temporal (a)synchrony of audiovisual speech (Stevenson, VanDerKlok, Pisoni, & James, 2011) found one subregion of STC to be more active for synchronous input whereas another subregion was more active for asynchronous input. However, our findings do not directly support this notion given that we did not identify any cluster in the STC that exhibited larger activity during incongruent McGurk trials than congruent stimulation.

---

**DISCUSSION**

In this fMRI study, we investigated neural activity differences between congruent and incongruent audiovisual stimulation in participants who consistently fused the incongruent audiovisual stimuli, that is, where auditory /aba/ and visual /aga/ were predominantly perceived as /ada/. We observed larger activity in STC for congruent audiovisual stimulation compared to incongruent McGurk stimuli, suggesting that STC cluster (p > .1). Thus, perceptual differences per se cannot explain the observed congruency effect in STC.

Furthermore, we observed more activity for congruent videos in the right angular gyrus. This region was part of a cluster that was removed by our inclusive masking procedure (see Figure 2), indicating that it was not reliably more active during task than during baseline. Furthermore, its anatomical location corresponded to one of the nodes of the “default mode network” (Esposito et al., 2006; Raichle et al., 2001).

---

**Table 1. Brain Regions Associated with Decreased Activity during McGurk Trials Compared to Congruent Audiovisual Stimulation (p < .001, Uncorrected)**

<table>
<thead>
<tr>
<th>Contrast</th>
<th>Anatomical Region</th>
<th>Hemisphere</th>
<th>Cluster Size</th>
<th>MNI Coordinates</th>
<th>T Value (df)</th>
</tr>
</thead>
<tbody>
<tr>
<td>AVcongruent &gt; McGurk</td>
<td>Superior temporal gyrus</td>
<td>Left</td>
<td>208 (Voxels)</td>
<td>x = −48 y = −24 z = 8</td>
<td>5.49 (22)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Right</td>
<td>109 (Voxels)</td>
<td>x = 58 y = −22 z = 16</td>
<td>4.75 (22)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>91 (Voxels)</td>
<td>x = 56 y = −2 z = 4</td>
<td>5.04 (22)</td>
</tr>
<tr>
<td></td>
<td>Angular gyrus</td>
<td>Right</td>
<td>85 (Voxels)</td>
<td>x = 58 y = −50 z = 22</td>
<td>5.14 (22)</td>
</tr>
</tbody>
</table>

The analysis was restricted to clusters that were activated by the task (inclusive mask threshold p < .001).

---

**Figure 2. Neuroimaging results.** Brain regions showing increased activity during congruent audiovisual (AV) stimulation (audiovisual /aba/ and /aga/) compared with McGurk trials (auditory /aba/, visual /aga/). Orange clusters were more active during the task compared to the implicit baseline, whereas blue clusters were not. A coronal (y = −22) and a horizontal slice (z = −4 left, z = 4 right) through the peaks of the clusters are shown. Results are thresholded at p < .001 at the voxel level.
Another factor possibly contributing to different outcomes regarding the response of STC to audiovisual congruence might be task effects. In this study, participants had to make a perceptual decision on all stimuli. This ensured that participants were actively processing all stimuli. Other studies have used passive listening (Skipper, van Wassenhove, Nusbaum, & Small, 2007) or detection of distractors (Nath & Beauchamp, 2012). It is possible that, when less attention is paid overall to the stimuli, McGurk stimuli might automatically attract attention to a stronger degree. Different task designs can, however, not account for different results entirely (Szychik et al., 2012).

**Different Shades of Audiovisual Congruency**

STC is sensitive to audiovisual congruencies of various kinds: spatial (i.e., sources of audiovisual inputs; e.g., Plank, Rosengarth, Song, Ellermeier, & Greenlee, 2012), temporal (i.e., onsets of sound and lip movements; e.g., Macaluso, George, Dolan, Spence, & Driver, 2004), or identity (i.e., perceived phonemes; e.g., Nath & Beauchamp, 2012; Szychik et al., 2012). We manipulated the congruency of identity in our study, but it is conceivable that differences in temporal congruency might also have contributed to the observed effect in STC, if there are differences in temporal asynchrony between McGurk stimuli and congruent videos. It should be noted that although even during congruent speech auditory and visual signals are not necessarily temporally synchronous (Schwartz & Savariaux, 2014; Chandrasekaran, Trubanova, Stillittano, Caplier, & Ghazanfar, 2009), and speech signals are still perceived as congruent if the temporal asynchrony does not exceed a certain maximum (Van Wassenhove, Grant, & Poeppel, 2007). The STC is generally susceptible to these temporal differences between congruent and incongruent speech (Stevenson, Altieri, Kim, Pisoni, & James, 2010).

It appears unlikely that temporal incongruence is responsible for the congruency effect that we found in STC for the following reasons. First, the temporal asynchrony of auditory and visual signals in the audiovisual stimuli was comparable for congruent and McGurk stimuli in our study. Second, close temporal coherence between auditory and visual signals is required to elicit the McGurk illusion (Van Wassenhove et al., 2007). This suggests that when incongruent audiovisual stimulation is merged by the brain it is not necessarily more asynchronous in time than congruent speech. Therefore, although STC is generally susceptible to temporal and spatial congruence, it appears likely that the congruency effect found for audiovisual congruent speech and the McGurk illusion in this study stems from incongruence in identity, namely congruent audiovisual /g/ and /b/ compared to conflicting visual /g/ and an auditory /b/. It should be noted that although the physical inputs of McGurk illusions are by definition incongruent, the merging process rendered the inputs perceptually congruent. Participants were unable to identify the inputs leading to the McGurk illusion, suggesting that they were unaware of the incongruent inputs. The congruency effect in STC therefore probably reflects the congruency of external stimulation rather than perceptual congruence.

**Candidate Mechanisms of Multisensory Integration**

Our finding that the STC favors congruent stimulation is in line with previous work on audiovisual congruency looking at letters and speech (Van Atteveldt et al., 2004, 2010) and the temporal congruency of spoken sentences and videos (Calvert et al., 2000). Although no firm conclusions can be drawn on the basis of these data alone, our results are in line with the notion that STC has a patchy organization that contains unisensory as well as multisensory neurons (Beauchamp, Argall, Bodurka, Duyn, & Martin, 2004). During congruent audiovisual stimulation, both unisensory neurons (representing unisensory auditory and visual components) and multisensory neurons (representing the combined audiovisual percept) are hypothesized to be active. These multisensory neurons are most strongly driven if they receive congruent input from both modalities, as indicated by single-unit recordings in STC of nonhuman primates (Dahl, Logothetis, & Kayser, 2010). However, during incongruent stimulation (i.e., McGurk stimuli), multisensory neurons may be driven less strongly, as they get conflicting input from different sensory modalities, culminating in reduced activity in STC for McGurk stimuli than congruent stimulation. At the same time, unisensory neurons may fire as strongly as during congruent stimulation because only the unisensory components but not the combined percept should matter for them. Thus, stronger activation of STC during congruent stimulation than incongruent stimulation could be explained by the existence of multisensory neurons in STC. If STC would only contain unisensory neurons, we would have expected no activity difference in STC for (in)congruent audiovisual stimulation. Under the hypothesis that the role of the STC is to merge information from the visual and auditory senses, it should also respond more strongly for McGurk illusions (when a merged percept is created) than for the same McGurk stimuli that do not elicit the illusion (when no merged percept is formed). Although we did not have enough nonillusion trials to investigate this hypothesis, this is indeed what has been observed previously, both within participants who did not perceive the illusion all the time (Szychik et al., 2012) and between participants who either were prone to the McGurk illusion or not (Nath & Beauchamp, 2012). STC is more active when participants merge the senses than when the senses cannot be merged, that is, when they perceive the McGurk illusion compared to when they do not.

The clusters we found in STC that responded more strongly for congruent than incongruent McGurk stimuli were partly overlapping with primary auditory cortex.
et al., 2000), it is conceivable that STC communicates inte-
cisions in the strength of the McGurk illusion (Nath &
English speakers are among other Western languages
and visual information streams do not match.

Conclusions

This study demonstrated a congruency effect in human
STC while controlling for individual differences in mul-
sensory integration. Although previous studies were
consistent in the direction of this effect, this study
sheds more light into the nature of this congruency effect,
namely, that STC vigorously responds to matching audio-
visual speech input and less to incongruent stimulation.
The selectivity to audiovisual congruency in STC does
not only apply to seeing mouth movements that match
the articulated auditory syllables, as shown in this study,
but also to concurrent reading and hearing of matching
letters (Van Atteveldt et al., 2010). Both of these studies
focus on speech-related audiovisual integration. It has to
be investigated whether STC plays a similar role in other
non-speech-related audiovisual processes like observing
actions of others. An electrophysiological study in mon-
keys suggests that the congruency effect indeed general-
izes to action observation by showing stronger activation
of STC for congruent than incongruent audiovisual stim-
ulation (Barracough, Xiao, Baker, Oram, & Perrett, 2005).

Furthermore, congruency can be adaptive to experi-
ence. Congruency effects in STC have been shown to de-
pend on the language context (Holloway, van Atteveldt,
Blomert, & Ansari, 2015). Similarly, the strength of the
McGurk illusion is affected by language. Although Dutch
and English speakers are among other Western languages
generally prone to the McGurk illusion (Van Wassenhove
et al., 2007), the effect seems to be weaker for Asian lan-
guages (Sekiyama & Tohkura, 1991). Congruency effects
in STC might be modulated by a dynamic interplay be-
tween long- and short-term experiences in the merging
of unsensory inputs into a multisensory percept.

Multisensory stimuli might be perceived as mismatch-
ing for some people whereas others perceive a combined
percept as nicely demonstrated by interindividual differ-
ences in the strength of the McGurk illusion (Nath &
Beauchamp, 2012). This study, furthermore, highlights
the importance of controlling for individual behavioral
differences in multisensory integration, which might con-
found the neuroimaging results. In summary, this study
supports the notion that the STC is a crucial binding site
for audiovisual integration that is most strongly driven by
congruent input.

Reprint requests should be sent to Claudia S. Lüttke, Donders
Institute for Brain, Cognition and Behaviour, Radboud University,
P.O. Box 9101, 6500 HB Nijmegen, the Netherlands, or via
e-mail: c.luettke@donders.ru.nl.

REFERENCES

of surprise with applications to attention. Neural Networks,
23, 649–666.

Barracough, N. E., Xiao, D., Baker, C. I., Oram, M. W., &
information by superior temporal sulcus neurons responsive
to the sight of actions. Journal of Cognitive Neuroscience,
17, 377–391.

Baum, S. H., Martin, R. C., Hamilton, A. C., & Beauchamp, M. S.
(2012). Multisensory speech perception without the left

Beauchamp, M. S., Argall, B. D., Bodurka, J., Duyn, J. H., &
Patchy organization within human STS multisensory cortex.
Nature Neuroscience, 7, 1190–1192.

transcranial magnetic stimulation reveals that the superior
temporal sulcus is a cortical locus of the McGurk effect.
Journal of Neuroscience, 30, 2414–2417.

Brugge, J. F., Volkov, I. O., Garell, P. C., Reale, R. A., &
auditory cortex on Heschl’s gyrus and on the lateral superior
temporal gyrus in humans. Journal of Neurophysiology,
90, 3750–3763.

Calvert, G. A., Brammer, M. J., Bullmore, E. T., Campbell, R.,
in sensory-specific cortices during crossmodal binding.
NeuroReport, 10, 2619–2623.

functional magnetic resonance imaging of crossmodal binding
in the human heteromodal cortex. Current Biology, 10,
649–657.

Chandrasekaran, C., Trusbano, A., Stillittano, S., Caplier, A., &

of visual responses in the superior temporal sulcus by
audio-visual congruency. Frontiers in Integrative
Neuroscience, 4, 10.

Den Ouden, H. E. M., Friston, K. J., Daw, N. D., McIntosh, A. R., &
Stephan, K. E. (2009). A dual role for prediction error in
associative learning. Cerebral Cortex, 19, 1175–1185.

Eickhoff, S. B., Paus, T., Caspers, S., Grosbras, M.-H., Evans,
activations to probabilistic cytoarchitectonic areas revisited.
Neuroimage, 36, 511–521.

Esposito, F., Bertolino, A., Scarabino, T., Latorre, V., Blasi, G.,
Popolizio, T., et al. (2006). Independent component model of
the default-mode brain function: Assessing the impact of

for improving auditory detection. Journal of the Acoustical


Schwartz, J. L., & Savariaux, C. (2014). No, there is no 150 ms lead of visual speech on auditory speech, but a range of audiovisual asynchronies varying from small audio lead to large audio lag. *PLoS Computational Biology, 10*, e1005743.


