Behavioral/Cognitive

The Motor Network Reduces Multisensory Illusory Perception

Takenobu Murakami,[1] Mitsunari Abe,[2] Winnugroho Wiratman,[1,6] Juri Fujiwara,[3] Masahiro Okamoto,[3] Tomomi Mizuochi-Endo,[7] Toshiki Iwabuchi,[7] Michiru Makuuchi,[7] Akira Yamashita,[1] Amanda Tiksnadi,[1,6] Fang-Yu Chang,[1] Hitoshi Kubo,[4] Nozomu Matsuda,[1] Shunsuke Kobayashi,[1] Satoshi Eifuku,[3] and Yoshikazu Ugawa[1,5]

[1]Department of Neurology, [2]Center for Neurological Disorders, [3]Department of Systems Neuroscience, [4]Advanced Clinical Research Center, Fukushima Global Medical Science Center, [5]Department of Neuro-regeneration, Fukushima Medical University, Fukushima, 960-1295, Japan, [6]Department of Neurology, Faculty of Medicine, University of Indonesia, Jakarta, 10430, Indonesia, and [7]Section of Neuropsychology, Research Institute of National Rehabilitation Center for Persons with Disabilities, Tokorozawa, 359-8555, Japan

Observing mouth movements has strikingly effects on the perception of speech. Any mismatch between sound and mouth movements will result in listeners perceiving illusory consonants (McGurk effect), whereas matching mouth movements assist with the correct recognition of speech sounds. Recent neuroimaging studies have yielded evidence that the motor areas are involved in speech processing, yet their contributions to multisensory illusion remain unclear. Using functional magnetic resonance imaging (fMRI) and transcranial magnetic stimulation (TMS) in an event-related design, we aimed to identify the functional roles of the motor network in the occurrence of multisensory illusion in female and male brains. fMRI showed bilateral activation of the inferior frontal gyrus (IFG) in audiovisually incongruent trials. Activity in the left IFG was negatively correlated with occurrence of the McGurk effect. The effective connectivity between the left IFG and the bilateral precentral gyri was stronger in incongruent than in congruent trials. The McGurk effect was reduced in incongruent trials by applying single-pulse TMS to motor cortex (M1) lip areas, indicating that TMS facilitates the left IFG-precentral motor network to reduce the McGurk effect. TMS of the M1 lip areas was effective in reducing the McGurk effect within the specific temporal range from 100 ms before to 200 ms after the auditory onset, and TMS of the M1 foot area did not influence the McGurk effect, suggesting topographical specificity. These results provide direct evidence that the motor network makes specific temporal and topographical contributions to the processing of multisensory integration of speech to avoid illusion.

Key words: functional magnetic resonance image; inferior frontal gyrus; McGurk effect; motor cortex; multisensory illusion; transcranial magnetic stimulation

---

Significance Statement

The human motor network, including the inferior frontal gyrus and primary motor cortex lip area, appears to be involved in speech perception, but the functional contribution to the McGurk effect is unknown. Functional magnetic resonance imaging revealed that activity in these areas of the motor network increased when the audiovisual stimuli were incongruent, and that the increased activity was negatively correlated with perception of the McGurk effect. Furthermore, applying transcranial magnetic stimulation to the motor areas reduced the McGurk effect. These two observations provide evidence that the motor network contributes to the avoidance of multisensory illusory perception.

Introduction

Human sensory systems develop and are tuned so that they can perform optimally in various environments. Multiple sensory inputs can be integrated to improve perception and decrease sensory uncertainty. Compared with unimodal stimuli, semantically

congruent and temporally matched multisensory inputs enhance the accuracy of perception and reduce the reaction time (RT) of recognition (Green and Angelaki, 2010), whereas incongruent stimuli have the opposite effects (Sekuler et al., 1997). The McGurk effect is one of the multisensory illusions elicited when phonetic sounds and lip movements are incongruent. For example, a combination of an auditory /Pa/ and a visual /Ka/ results in the perception of a different syllable, /Ta/ (McGurk and Mac-Donald, 1976). The posterior part of the superior temporal sulcus (STS) is an important candidate region for the audiovisual integration of congruent stimuli (Calvert et al., 2000; Beauchamp et al., 2004) and even of incongruent McGurk syllables (Nath and Beauchamp, 2012; Matchin et al., 2014). On the other hand, neuroimaging data have shown that the motor areas—including the inferior frontal gyrus (IFG), dorsal part of the premotor cortex (dPMC), and primary motor cortex (M1)—are activated during audiovisually incongruent speech, including McGurk stimuli, compared with congruent speech (Bushara et al., 2001; Miller and D'Esposito, 2005; Ojanen et al., 2005; Skipper et al., 2007; Benoit et al., 2010).

The recruitment of motor areas during audiovisual speech can be explained by the motor theory of speech perception, whereby speech sounds are recognized based on the motor representations that underlie speech gestures (Liberman and Mattingly, 1985). This motor theory is closely linked to the framework of a mirror neuron system, which plays a role in mapping sensory inputs onto matching motor representations for understanding actions (Rizzolatti and Arbib, 1998; Corballis, 2010). However, the functional contributions of the motor areas to multisensory illusory perception remain to be determined.

Transcranial magnetic stimulation (TMS) is widely used in many neuroscience fields, including speech studies. The application of TMS with an event-related design temporally disrupts or facilitates the activity of the stimulated network, which should result in a deterioration or improvement of the behavioral performance if this network is relevant to the performed task (Murakami et al., 2013). However, the underlying mechanisms are not well understood and are explained poorly when invoking simple mechanisms (Schwarzkopf et al., 2011; Perini et al., 2012). Beauchamp et al. (2010) reported on a unique event-related TMS experiment using a McGurk task. Participants had to identify syllables from audiovisually congruent or incongruent McGurk stimuli after event-related single-pulse TMS over the left STS versus the control site and no TMS. TMS of the left STS selectively reduced the occurrence of the McGurk effect compared with stimulation of the control site and the no-TMS condition. Beauchamp et al. (2010) showed that TMS creates a temporary virtual lesion in the STS and leads to a reduction of the McGurk effect. However, TMS does not necessarily induce inhibitory effects, and it is also necessary to determine how the stimulated network is involved in the behavioral performance.

To investigate the functional role of the motor network in multisensory illusory speech perception, we conducted two experiments using audiovisual McGurk stimuli with an event-related design. First, functional magnetic resonance imaging (fMRI) was performed to identify activations of the motor areas with McGurk stimuli. Correlation analysis and psychophysiological interaction (PPI) analysis were used to clarify the motor areas related to the McGurk effect. Second, we applied single-pulse TMS over the M1 lip area and found temporally and topographically specific reduction of McGurk illusion susceptibility (IS). These approaches showed that the motor network is involved in the occurrence of multisensory illusion.

Materials and Methods

Participants

The study participants comprised 26 healthy right-handed subjects (including 10 females) aged 21.4 ± 4.0 years (mean ± SD) with a score on the Edinburgh Handedness Inventory of 85.8 ± 19.6% (Oldfield, 1971). Written informed consent was obtained from all subjects before their participation. The study was approved by the ethics committee of Fukushima Medical University (approval no. 2008) and conformed to the latest version of the Declaration of Helsinki.

Behavioral paradigm

The applied audiovisual stimuli were recordings of a native Japanese male speaker vocalizing the sounds /Ba/, /Pa/, /Ga/, and /Ka/. The total length of each video clip (frame size: 720 × 480 pixels; frame rate: 29 fps) was 2500 ms, and the duration of each auditory syllable was ~300 ms. Visual lip movements began 60 ms before the onset of the auditory stimulation. Audiovisually incongruent (auditory /Ba//Pa/ and visual /Ga//Ka/) and congruent (both /Ba//Pa/) perceptual tasks were presented using a stimulus delivery control program (Presentation, Neurobehavioral Systems). In the fMRI experiment, auditory stimuli were delivered at ~70 dB SPL via headphones compatible with fMRI, and visual stimuli were presented on a computer screen via a mirror. In the TMS experiment, auditory stimuli were presented via noise-isolating in-ear headphones at ~70 dB SPL, and visual materials were presented on a PC screen positioned 1 m directly in front of the participant. The subjects were instructed to focus on a cross cue for 1 s before the task and then to watch the mouth movements while listening to the speaker. Three written syllables were then presented on a screen, and subjects chose the syllable that they had heard by pressing one of three response buttons; the three possible syllables were those representing the auditory, visual, and McGurk perceptions (Fig. 1A). Twenty-four trials of incongruent and congruent tasks were presented, with the tasks delivered in random order and separated by 8–10 s intertrial intervals.

Experiment 1: functional localization related to audiovisual illusion in an event-related fMRI experiment

All 26 subjects participated in Experiment 1.

MRI recording. fMRI and high-resolution structural T1-weighted image data were acquired on a 3.0-tesla Siemens Biograph mMR System equipped with a 12-channel head coil. An fMRI experiment for audiovisual trials was performed with an event-related design using a blood oxygen level-dependent (BOLD) effect-sensitive gradient echo planar imaging (EPI) sequence covering the entire brain [repetition time (TR) = 3.0 s, echo time (TE) = 30 ms, field-of-view (FOV) = 192 mm, flip angle = 90°, 64 × 64 matrix, 45 slices, slice thickness = 3.0 mm, and voxel size = 3.0 × 3.0 × 3.0 mm]. To confirm the localization of the M1 lip areas and the left foot area, other fMRI experiments were performed in a block design. The subjects had to say /Ba//Pa/ when the written syllables appeared on a computer screen for 30 s. In an additional condition, the word "rest" appeared on the screen for 30 s, during which the subjects rested. Each task was performed six times alternatively. The subjects also performed ballistic movements of the right foot and rested alternatively (6 times for 30 s each) according to written instructions. Coregistration with the functional data was performed by acquiring a high-resolution structural T1-weighted image using an MRI machine(TR = 1800 ms, TE = 1.99 ms, FOV = 250 mm, flip angle = 9°, 256 × 256 matrix, 176 slices, slice thickness = 1 mm, and voxel size = 1.0 × 1.0 × 1.0 mm).

fMRI data analysis. fMRI data were processed in a standard manner using SPM12 software (Wellcome Department of Imaging Neuroscience, London, UK). EPI volumes were realigned, spatially normalized to the Montreal Neurological Institutes (MNI) space, and smoothed using a Gaussian filter with a full-width at half-maximum of 8 mm. The preprocessed functional images were analyzed using the standard general linear model approach. Beta maps of each audiovisual task and a simple t-contrast map of incongruent versus congruent tasks were calculated for each subject. Group data from the β images were analyzed using simple one-sample t tests. Group data for saying /Ba//Pa/ versus a resting condition and for moving a foot versus a resting condition were also analyzed
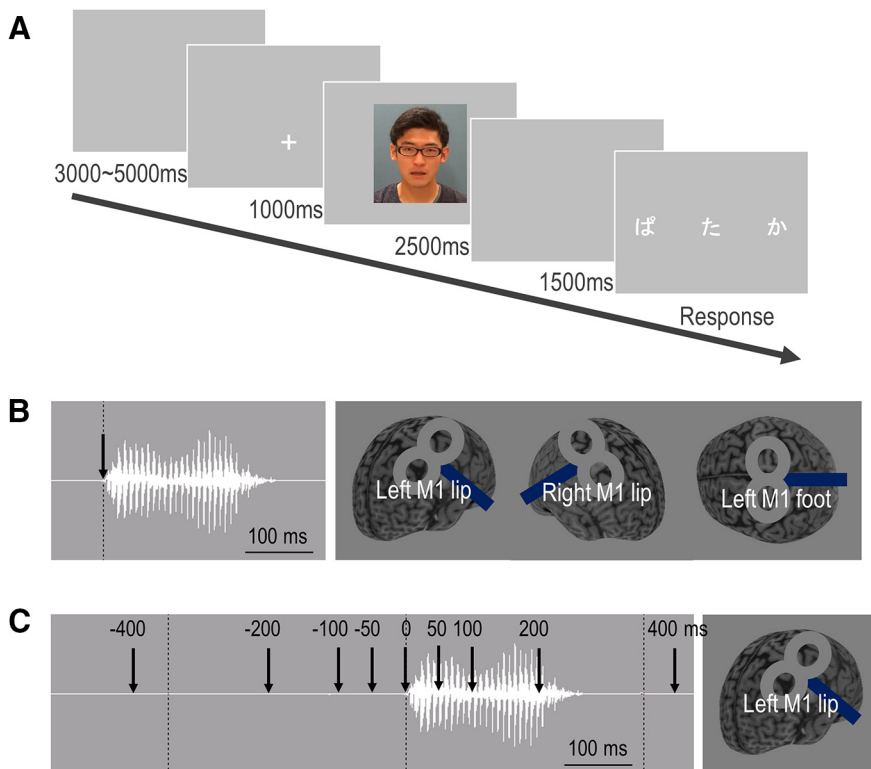
A

B

C

Figure 1. Experimental protocol. *A*, Subjects were presented with a McGurk illusory task interspersed within a series of audiovisually congruent trials with an event-related design. Subjects had to choose the syllable they perceived from three possibilities. *B*, Event-related TMS study in Experiment 2-1. Subjects received single-pulse TMS at the onset of an auditory syllable. TMS was applied over bilateral M1 lip areas and the M1 foot area using a TMS neuronavigation system guided by fMRI. *C*, Event-related TMS study in Experiment 2-2. Single-pulse TMS was applied over the left M1 lip area in every trial at one of nine times, from 400, 200, 100, or 50 ms before or 0, 50, 100, 200, or 400 ms after the auditory onset.

with one-sample *t* tests. Clusters were considered significant if they consisted of at least 100 contiguous voxels that passed the threshold of $p < 0.05$ with familywise error (FWE) correction in the voxel-based analyses.

Behavioral data analysis. The susceptibility to McGurk effect (i.e., IS) was calculated as the proportion of McGurk effect occurrences relative to the total stimuli in each trial. RT was defined as the time required for a subject to respond to the three syllables presented on a screen. IS and RT were calculated in both the audiovisually incongruent and congruent trials, and they were analyzed using paired *t* tests.

Correlation analysis. The correlation between task-related activity in the left IFG and behavioral data was analyzed using contrast maps of incongruent versus congruent trials. Individual differences in IS and RT between incongruent trials and congruent trials (delta IS and delta RT, respectively) were calculated and used as covariates. The significance threshold was set at 100 contiguous voxels for which $p < 0.05$ with FWE small-volume correction in a cluster of the left IFG identified in the activation map of incongruent versus congruent tasks, because the left IFG is a candidate area involved in the processing of multisensory illusory perception (Miller and D'Esposito, 2005; Ojanen et al., 2005).

PPI analysis. PPI analysis was applied to the functional connectivity of the brain regions associated with modulation of the McGurk effects according to motor areas (Friston et al., 1997). At the individual level, spheres with radii of 8 mm were created around the IFG and STS in the left hemisphere derived from the earlier one-sample *t* tests and correlation analysis. The time series of the BOLD response for each subject was computed using the first eigenvariate from all time series of the voxels in the spheres. The BOLD time series for each subject was then deconvolved to estimate a neuronal time series for the seed using the PPI deconvolution parameter. The PPI regressor was calculated as the element-by-element product of the neuronal time series and a vector coding for the main effect of the task. The P regressor represented the contrast of incongruent versus congruent tasks (the psychological variable), whereas the Y

regressor showed the seed neuronal time series (the physiological variable). These regressors were reconvolved using the canonical hemodynamic response function. Subject-specific PPI models were run, and images were produced of the changes in connectivity with the source presented as the contrast between incongruent and congruent tasks. These first-level contrast images were entered into a second-level general linear model to assess target regions for effective connectivity with the source. The targets were chosen in the left precentral gyrus [MNI coordinates $(x, y, z) = (-50, 0, 47)$] right precentral gyrus (49, 0, 51), left STS (-62, -42, 10), and left IFG (-48, 25, 19), because these are associated with audiovisual speech perception (Ojanen et al., 2005; Skipper et al., 2005; Hocking and Price, 2008). The target regions were selected in the spheres (radius = 8 mm) centered at these coordinates. The statistical threshold was set at 20 contiguous voxels for which $p < 0.05$ with FWE small-volume correction.

Experiment 2: effects of TMS on audiovisual illusion in an event-related TMS experiment
Twenty-four of the 26 subjects participated in Experiment 2.

Transcranial magnetic stimulation. TMS was delivered using a magnetic stimulator connected to a figure-of-eight coil with external loops having a diameter of 70 mm (Magstim 200, Magstim). The magnetic stimulus had a monophasic waveform. The stimulating coil was placed tangentially to the scalp with the handle pointing backwards and laterally 45° from the anterior–posterior axis. The TMS intensity was set at the motor threshold, which was defined as the minimum stimulus intensity that elicited a small motor-evoked potential (MEP) with a peak-to-peak amplitude of >50 μV from the orbicularis oris muscles in at least 5 of 10 consecutive trials. The stimulus intensity was indicated as a percentage of the maximum stimulator output. Because it is difficult to keep the orbicularis oris muscles fully at rest, the participants were instructed about how to maintain electromyogram quiescence for $\sim$10 min. This was achieved by providing the participants with continuous high-gain visual (50 μV/division) and auditory feedback of the electromyogram activity (Murakami et al., 2011).

fMRI-guided TMS neuronavigation. Stimulation sites of the bilateral M1 lip areas and the left M1 foot area were identified on the scalp of each subject using an fMRI-guided TMS neuronavigation system (Brainsight, Rogue Research). Individual fMRI-guided TMS neuronavigation is currently the most precise approach for determining the target regions given the high interindividual variability in the relationship between the spatial location of brain function and anatomical landmarks (Sack et al., 2009). The 3D high-resolution structural T1-weighted image of an individual was imported, and this image was standardized to the Talairach coordinate system by defining the anterior-commissure–posterior-commissure line and the falx cerebri. Four skull landmarks of each subject (nasion, tip of nose, and bilateral preauricular points) were fitted to those on the 3D image. Errors of up to 3 mm were allowed between the scalp and the image. The MNI coordinates of individual activated regions in fMRI of the M1 lip areas and the M1 foot area were automatically transformed into Talairach coordinates using the Brainsight software, and the stimulating coil was visually navigated to the stimulation target and kept there with the aid of real-time feedback of the coil position throughout each TMS experiment. The individual coordinates for TMS targeting were determined as the maximum individual fMRI activations in close proximity to the bilateral M1 lip areas and the left M1 foot area as
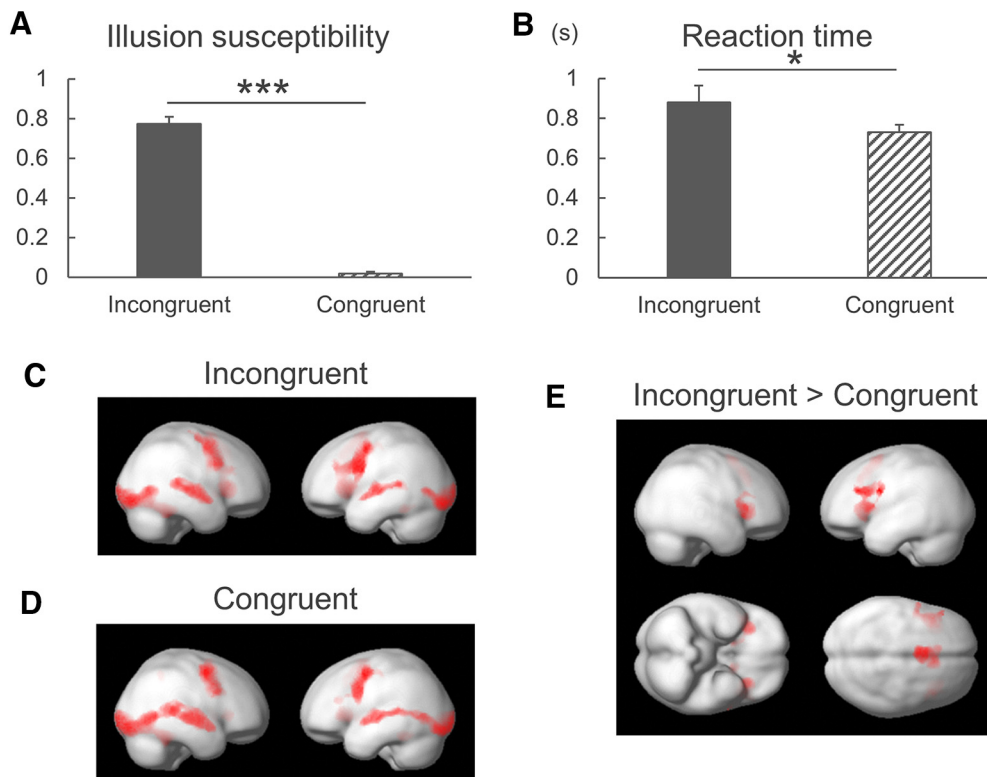
Figure 2. Behavioral and fMRI results. ***A***, The rate of McGurk IS was higher in audiovisually incongruent trials than in congruent trials. ***B***, The reaction time was longer in incongruent trials than in congruent trials. *$p < 0.05$, ***$p < 0.001$. Data are mean and SEM values. ***C***, Audiovisually incongruent trials and (***D***) congruent trials widely activated audiovisual areas and the frontal lobes, including the dPMC and the IFG in both hemispheres. ***E***, Contrast maps of incongruent versus congruent trials showed stronger activations in the bilateral IFG, anterior cingulate gyrus, and putamen. Activations were considered positive when there were at least 100 contiguous voxels that passed the threshold of $p < 0.05$ with familywise error correction.

defined in the group fMRI data, to take into account interindividual differences (Murakami et al., 2012, 2015).

Event-related single-pulse TMS. The delivery of single-pulse TMS and audiovisual stimuli was controlled using the Presentation program. In Experiment 2–1, each subject participated in three sessions; two with the TMS targeting the M1 lip areas bilaterally and one with the TMS over the left M1 foot area. The order of TMS sites was set randomly. Single-pulse TMS was applied at the onset of the auditory syllable. In each session, each run consisted of 12 trials of randomly intermixed audiovisually incongruent and congruent tasks with and without TMS, resulting in a total of 48 trials. The data obtained in the 12 trials were averaged in each TMS condition, and those of the no-TMS condition (total 36 trials) were averaged and used for analyzing the no-TMS condition. To examine the consistency of TMS effects within subjects, we performed a test–retest experiment in the condition of TMS over the left M1 lip area and the no-TMS condition in a random order. The second set of recordings was made at least 10 d after the first set.

In Experiment 2-2, single-pulse TMS was applied over the M1 lip area on the left hemisphere in each trial for a total of nine times: 400, 200, 100, or 50 ms before or 0, 50, 100, 200, or 400 ms after the auditory onset. Each run of 12 trials of audiovisually incongruent tasks were delivered with and without TMS in a random order, to produce a total of 120 trials.

In Experiment 2-1, McGurk IS and RT were compared using a two-way repeated-measures ANOVA (rmANOVA) with trial (2 levels: incongruent and congruent) and stimulus site (4 levels: left M1 lip area, right M1 lip area, left M1 foot area, and no TMS) as the within-subject factors. To clarify whether TMS shifted the illusory perception to auditory or visual perception in incongruent trials, we analyzed the rates of non-McGurk perception (auditory /Ba//Pa/ and visual /Ga//Ka/) chosen from the three syllables. An additional two-way rmANOVA was performed with audiovisual perception (2 levels: auditory- and visual-based responses) and stimulus site (4 levels: left M1 lip area, right M1 lip area, left M1 foot area, and no-TMS) as the within-subject factors. We calculated

Pearson's correlation coefficients for the data in the test–retest experiment to confirm the reproducibility of the McGurk effect.

In Experiment 2-2, IS and RT were compared using one-way rmANOVA with stimulus timing (10 levels: −400, −200, −100, −50, 0, 50, 100, 200, and 400 ms from the onset of the auditory syllable, and no TMS) as the within-subject factor. An additional two-way rmANOVA with audiovisual perception and stimulus timing as the within-subject factors was performed in both experiments to clarify how TMS interfered with illusory perception in incongruent trials. When a significant main effect was detected, *post hoc* analyses were performed using Fisher's protected least-significant-difference test.

In all of the above-described statistical analyses, which were performed using IBM SPSS Statistics v22.0, significance was assumed if $p < 0.05$.

Results
Experiment 1: results from the event-related fMRI experiment

Paired t tests showed that the behavioral parameters of the rate of IS was higher ($T = 19.041$, $p < 0.001$; Fig. 2A) and RT was longer ($T = 2.722$, $p = 0.012$; Fig. 2B) in audiovisually incongruent trials than in congruent trials.

A voxel-based analysis was first applied over the whole brain to explore regions associated with audiovisually congruent and incongruent conditions. Brain activations were widely distributed in audiovisual areas, the dPMC and IFG, bilaterally in both the audiovisually incongruent (Fig. 2C) and congruent (Fig. 2D) trials, which is consistent with the dual speech processing model (Hickok and Poeppel, 2007). The contrast maps of incongruent versus congruent trials revealed stronger activations of the bilateral IFG, anterior cingulate gyrus, and putamen (Fig. 2E; Table 1). In contrast, the activities of the STS, inferior temporal lobes,

Table 1. MNI coordinates of clusters activated by the contrast incongruent versus congruent trials

Region	Hemisphere	x	y	z	Z-score	k	P
Incongruent > congruent							
IFG	Right	30	30	6	5.26	992	<0.001
IFG	Left	−42	14	10	4.9	1445	<0.001
	Left	−38	16	24	4.53		
ACG	Left	−6	6	60	4.62	1263	<0.001
	Right	10	22	40	4.5		
Putamen	Left	−16	8	6	3.97	197	0.035
Incongruent < congruent							
ITG	Left	−58	−24	−18	4.4	489	<0.001
PCG	Left	−2	−66	20	4.22	281	0.007
IPL	Left	−62	−40	28	4.12	196	0.035
STS	Left	−50	−62	18	4.05	602	<0.001
MTG	Right	42	−72	26	4.02	393	0.001

ACG, Anterior cingulate gyrus; PCG, precentral gyrus; MTG, Middle temporal gyrus.

$N = 26$, cluster threshold = 100 voxels, $p < 0.05$ FWE-corrected.

and inferior parietal lobules in the left hemisphere were lower in incongruent trials than in congruent trials (Table 1). These results suggest that the IFG and STS in the left hemisphere are associated with processing of audiovisual illusory perception, but with different activation patterns: the left IFG is activated more in the audiovisually incongruent condition, whereas the left STS is activated more in the congruent condition.

To clarify the functional role of the left IFG in the audiovisual illusion, we analyzed the correlation between the task-related brain responses and the behavioral McGurk effect (delta IS and delta RT). The activity in the left IFG cluster (−38, 16, 24) was negatively correlated with delta IS (Fig. 3; $p < 0.05$ with FWE small-volume correction), which suggests that activation of the left IFG contributes to a reduction of the McGurk effect. However, delta RT was not significantly correlated with the IFG in the left hemisphere ($p > 0.05$, uncorrected).

We performed PPI analyses to examine whether the connectivity between the seeds and the target regions changed with the trial congruency. The coordinates of the left IFG (−38, 16, 24) and the left STS (−50, −62, 18) were used as seeds (Table 1). The PPI analysis revealed the influences of trial congruency on the connectivity: the effective connectivity between the left IFG and the bilateral precentral gyri (left: −54, 0, 50; right: 50, 4, 50) was stronger in incongruent trials than in congruent trials (Fig. 4; $p < 0.05$ with FWE small-volume correction), with the same relationship found between the left IFG and the left STS (−68, −38, 10) ($p < 0.05$ with FWE small-volume correction). In contrast, the PPI analysis produced the opposite result for the effective connectivity between the left STS and the left precentral gyrus (−54, 0, 42): the effective connectivity between the two regions of interest was stronger in congruent trials than in incongruent trials ($p < 0.05$ with FWE small-volume correction). However, there was no greater effective connectivity of the left STS with the precentral gyri nor the left IFG in incongruent trials than in congruent trials ($p > 0.05$, uncorrected).

Experiment 2: results from the event-related TMS experiment

The event-related TMS experiment aimed to extend the findings obtained in Experiment 1 by assessing the temporal and topographical causality of the motor areas in the McGurk effect.

Experiment 2-1: topographically specific modulation of the McGurk effect

Nineteen subjects participated in Experiment 2-1. Individual MNI coordinates of TMS sites were shown in Figure 5A and

Table 2. For IS, the two-way rmANOVA showed significant main effects of trial ($F_{(1,18)} = 273.132$, $p < 0.001$, $\eta^2 = 0.938$) and stimulus site ($F_{(3,54)} = 3.333$, $p = 0.026$, $\eta^2 = 0.156$), and a significant interaction ($F_{(3,54)} = 7.754$, $p < 0.001$, $\eta^2 = 0.301$). Illusion was perceived mostly in the incongruent trials, and TMS over the left and right M1 lip areas reduced IS significantly compared with TMS over the M1 foot area and the no-TMS condition. The topographically specific effects of TMS on IS reduction are supported by the findings of the *post hoc* paired t tests, with single-pulse TMS of the M1 lip areas bilaterally reducing IS in audiovisually incongruent trials compared with TMS over the M1 foot area (left M1 lip area, $p = 0.006$; right M1 lip area, $p = 0.013$) and the no-TMS condition (left M1 lip area, $p = 0.003$; right M1 lip area, $p = 0.002$; Fig. 5B). The magnitude of the IS reduction induced by TMS did not differ significantly between the left and right M1 lip areas ($p = 0.823$). Control stimulation to the left M1 foot area did not induce a significant change in IS ($p = 0.526$).

In audiovisually congruent trials, there were no significant differences in the modulation of IS between the conditions with and without TMS (all $p > 0.1$; Fig. 5B). We also measured the rates of non-McGurk perception for audiovisually incongruent stimuli to investigate how TMS shifted McGurk illusory perception to auditory or visual perception. Two-way rmANOVA revealed a significant main effect of stimulus site ($F_{(3,54)} = 6.895$, $p = 0.001$, $\eta^2 = 0.277$) but not of audiovisual perception ($F_{(1,18)} = 0.864$, $p = 0.365$, $\eta^2 = 0.046$), and no significant interaction ($F_{(3,54)} = 0.555$, $p = 0.647$, $\eta^2 = 0.030$). These results indicate that TMS did not disrupt the processing of auditory or visual perception, whereas it did block the occurrence of illusory perception.

To examine the reproducibility of TMS and the McGurk effect within subjects, we investigated 14 subjects in retest experiment. Although IS varied across the subjects, as also reported previously (Benoit et al., 2010; Nath and Beauchamp, 2012), we found a small amount of within-subject variability when TMS was applied over the left M1 lip area ($r = 0.870$, $p < 0.001$; Fig. 5C) and when TMS was not applied ($r = 0.616$, $p = 0.019$; Fig. 5D). These results confirm the reliability of the psychological measure of IS and the physiological effect of TMS over the M1 lip area on the McGurk effect.

For the RT, the rmANOVA revealed a significant main effect of trial ($F_{(1,18)} = 24.219$, $p < 0.001$, $\eta^2 = 0.574$) but not of TMS, and no interaction (both $F < 1.7$, $p > 0.1$, $\eta^2 < 0.1$), indicating that the RT is longer in audiovisually incongruent trials than in congruent trials, but also that single-pulse TMS has less impact on the RT.

Experiment 2-2: temporal modulation of the McGurk effect

Twelve subjects participated in Experiment 2-2, seven of whom had also participated in Experiment 2-1. Individual MNI coordinates of TMS sites were demonstrated in Figure 6A and Table 3. For IS, the one-way rmANOVA showed a significant main effect of time ($F_{(9,99)} = 4.587$, $p < 0.001$, $\eta^2 = 0.294$). The *post hoc* tests revealed that IS was significantly reduced by the application of single-pulse TMS from 100 ms before to 200 ms after the onset of the audiovisual stimuli ($p < 0.05$; Fig. 6B). When assessing the rates of auditory or visual perception for audiovisually incongruent stimuli, two-way rmANOVA showed a significant main effect of time ($F_{(9,99)} = 4.835$, $p < 0.001$, $\eta^2 = 0.305$) and an interaction ($F_{(9,99)} = 2.537$, $p = 0.012$, $\eta^2 = 0.187$), but no significant main effect of audiovisual perception ($F_{(1,11)} = 1.026$, $p = 0.333$, $\eta^2 = 0.085$). However, a *post hoc* paired t test revealed no differences
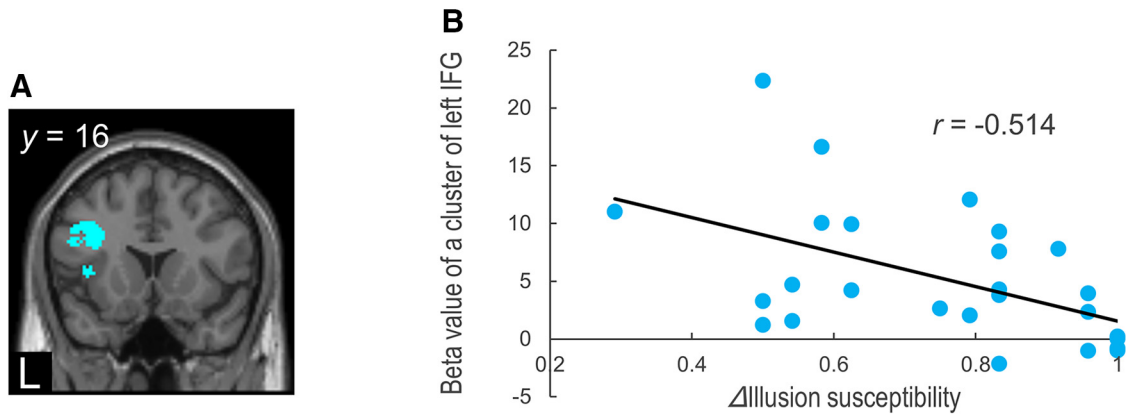
Figure 3. Correlation of fMRI data with behavioral results. *A*, A cluster in the left IFG showed a negative correlation with differences in McGurk IS between incongruent and congruent trials. *B*, Beta weights of the left IFG for the contrast of incongruent versus congruent trials were negatively correlated with differences in McGurk IS.
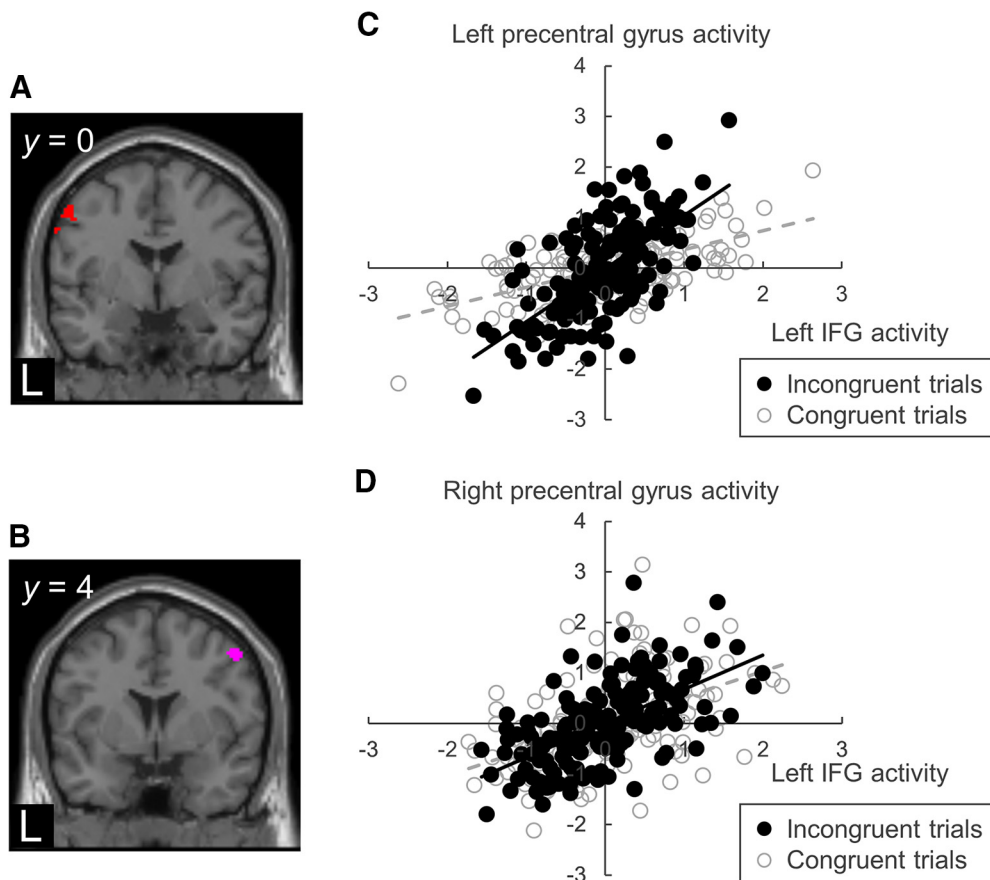

Figure 4. Psychophysiological interaction between the left IFG and the bilateral precentral gyri. The effective connectivity showed that the left IFG had positive connections with the (*A*) left precentral and (*B*) right precentral gyri in incongruent trials compared with in congruent trials. Data are from representative subjects demonstrating the correlation between the left IFG activity (*x*-axis) and (*C*) left precentral and (*D*) right precentral gyri activity (*y*-axis) during audiovisual tasks. Regression lines are black solid lines for the incongruent trials and gray dashed lines for the congruent trials.

between the rates of auditory and visual perceptions at any time point ($p > 0.05$), supporting the notion that TMS decreased the appearance of illusory perception but did not prevent the processing of auditory or visual perception at any time point.

For the RT, the one-way rmANOVA revealed a significant main effect of time ($F_{(9,99)} = 2.543$, $p = 0.011$, $\eta^2 = 0.188$). However, the *post hoc* analysis did not reveal any differences between the RT for any timing of single-pulse TMS and the RT without TMS ($p > 0.05$).

Discussion

This study provided empiric evidence for the contribution of the motor network to audiovisual illusions based on the effective connectivity (Experiment 1) and the causal connectivity (Experiment 2). Activation of the left IFG was associated with a reduction of IS. The effective connectivity between the left IFG and the bilateral precentral gyri was stronger in incongruent trials than in congruent trials. We also found that time-locked single-pulse
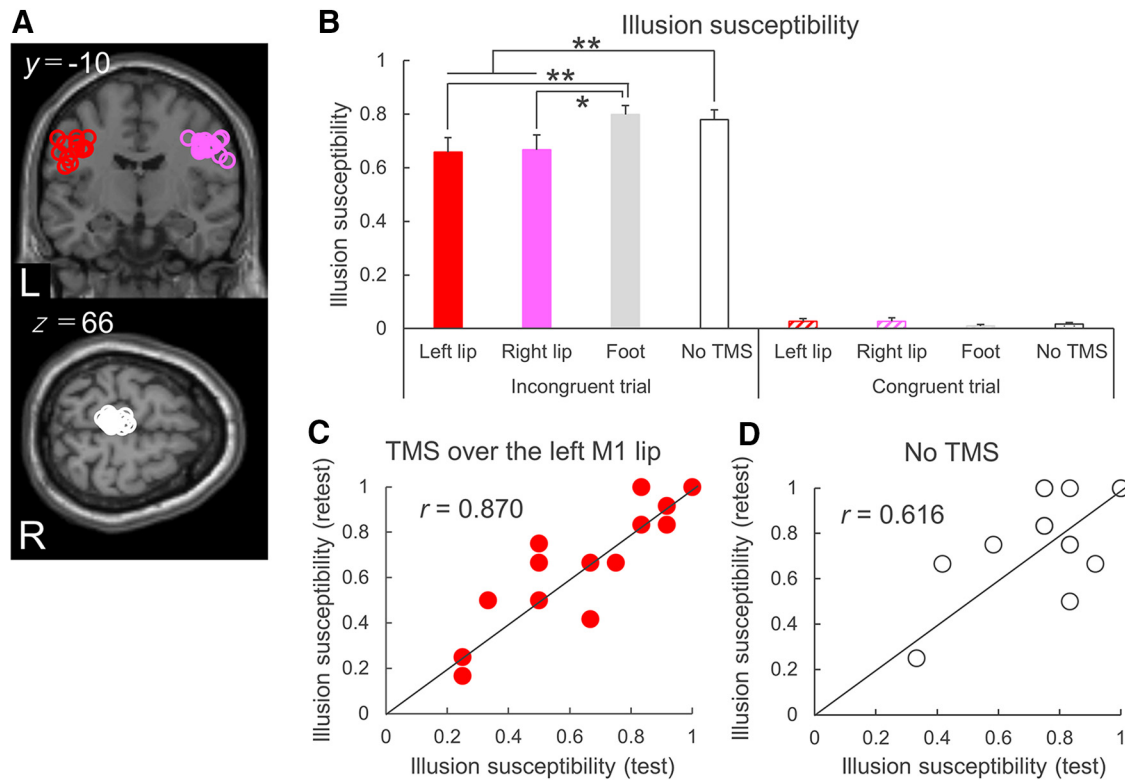
Figure 5. Topographical TMS effects on McGurk IS. *A*, Individual stimulation sites over bilateral M1 lip areas and the left M1 foot area are overlaid on brain images presented according to the standard MNI space. Overlap resulted in not all stimulation sites being visible. *B*, In audiovisually incongruent trials, McGurk IS was lower for single-pulse TMS over the bilateral M1 lip areas than for TMS of the M1 foot area and in the no-TMS condition. There were no significant differences in IS in congruent trials. *C*, Test–retest analysis in the condition of TMS over the left M1 lip area. Correlation analysis demonstrated a strong positive correlation of IS between the test and retest trials in TMS over the left lip area and (*D*) in the no-TMS condition. *$p < 0.05$, **$p < 0.01$. Data are mean and SEM values.

Table 2. MNI coordinates of individual stimulation sites in the Experiment 2-1

	Left M1 lip area			Right M1 lip area			Left M1 foot area		
	x	*y*	*z*	*x*	*y*	*z*	*x*	*y*	*z*
A	−48	−8	24	42	−10	40	−4	−30	66
B	−52	−12	38	56	−6	40	−4	−36	70
C	−52	−2	22	44	−12	34	−4	−22	64
D	−44	−10	34	50	−10	38	−8	−34	64
E	−52	−2	22	44	−14	34	−8	−28	66
F	−38	−14	34	48	−6	36	−8	−40	64
G	−36	−18	42	56	−10	30	−2	−26	78
H	−40	−12	34	62	−6	26	−6	−38	54
I	−58	−8	42	48	−6	34	−12	−36	72
J	−52	−12	38	50	−6	30	−6	−24	66
K	−56	−16	32	44	−14	36	−10	−24	72
L	−44	−14	42	44	−14	38	−10	−36	68
M	−50	−4	28	58	2	42	−4	−22	62
N	−44	−12	38	46	−8	40	−6	−38	72
O	−38	−14	34	44	−6	34	−6	−36	64
P	−38	−14	34	42	−12	34	−4	−34	68
Q	−40	−16	36	42	−10	32	−4	−34	70
R	−40	−14	34	36	−16	42	−10	−34	78
S	−44	−10	38	42	−12	34	−8	−34	72
Average ± SD	−45.6 ± 6.8	−11.2 ± 4.6	34.0 ± 6.2	47.3 ± 6.6	−9.3 ± 4.3	35.5 ± 4.3	−6.5 ± 2.7	−31.9 ± 5.8	67.9 ± 5.7

TMS over the M1 lip areas reduced IS in incongruent trials within a specific time range, indicating the causal influence of TMS on the McGurk effect via its interaction with the motor network processing. The motor network (including the left IFG) appears to perform specific modulation of multisensory perception in a task-specific manner.

Several studies have suggested that the STS is a critical region for multisensory integration of audiovisual speech perception (Calvert et al., 2000; Beauchamp et al., 2004; Matchin et al., 2014). This is supported by anatomical evidence that the STS is located between visual and auditory sensory cortices and exhibits strong connectivity (Seltzer and Pandya, 1978). McGurk perceiv-
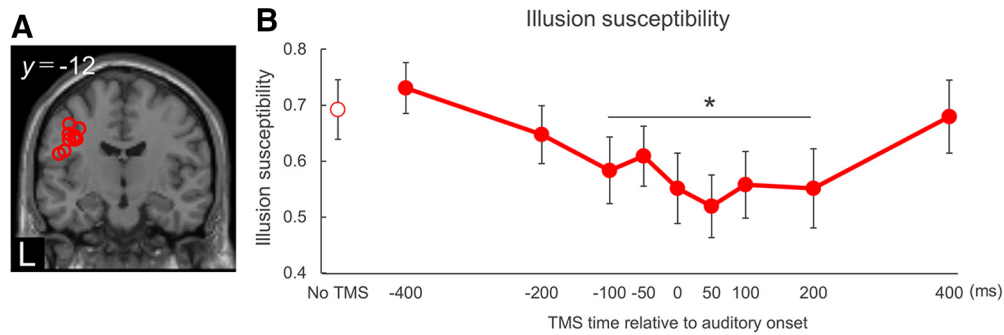
Figure 6. Temporal TMS effects on McGurk IS. ***A***, Individual stimulation sites over the left M1 lip area are overlaid on a standard MNI image of the brain. Overlap resulted in not all stimulation sites being visible. ***B***, Single-pulse TMS of the left M1 lip area reduced McGurk IS from 100 ms before to 200 ms after the onset of audiovisual stimuli compared with the no-TMS condition. *$p < 0.05$. Data are mean and SEM values.

Table 3. MNI coordinates of individual stimulation sites in the Experiment 2-2

	Left M1 lip area		
	x	y	z
A	−48	−8	24
B	−52	−2	22
C	−44	−10	34
D	−38	−14	34
E	−36	−18	42
F	−40	−12	34
G	−44	−12	38
H	−40	−14	36
I	−44	−10	38
J	−40	−16	36
K	−40	−14	36
L	−44	−12	46
Average ± SD	−43.7 ± 5.9	−11.3 ± 4.7	33.8 ± 6.1

ers show stronger STS activation compared with nonperceivers (Nath and Beauchamp, 2012). In addition, event-related single-pulse TMS over the left STS decreased the rate of McGurk effect, indicating that the STS plays a key role in integrating multisensory stimuli into a combined perception (Beauchamp et al., 2010).

Other neuroimaging studies have highlighted that the motor speech regions are also involved in multisensory processing (Miller and D'Esposito, 2005; Ojanen et al., 2005; Skipper et al., 2007; Benoit et al., 2010). Both the IFG and STS are activated during audiovisual matching and conflicting stimuli, and that the left IFG is more active in conflicting perception while the STS is not (Ojanen et al., 2005). The activity in the left IFG increases during unfused audiovisual stimuli but decreases during fused perception, whereas the left STS shows greater activity for fused stimuli than for unfused stimuli (Miller and D'Esposito, 2005). These results suggest that the roles of the left IFG and the STS in audiovisual integration may be separable, with the left IFG and STS relating to unmatched and matched audiovisual processing, respectively. The fMRI experiments in the present study also showed higher activations of the frontal areas (including the left IFG) in the contrast maps of audiovisually incongruent versus congruent trials, whereas the left STS was activated in the contrast of congruent versus incongruent trials. Furthermore, the activity of the left IFG activity was negatively correlated with the occurrence of the McGurk effect. This supports the hypothesis that the left IFG constitutes a part of the frontal lobe network that is generally responsible for modulating incompatible multisensory inputs (Novick et al., 2005).

The motor system is involved in the processing of audiovisual perception, with audiovisual stimulation activating the motor network over a large area, including the left IFG, bilateral dPMC, left ventral part of premotor cortex, and left M1 (Skipper et al., 2007). The precentral gyrus was activated more by audiovisually incongruent stimuli than by congruent stimuli, and that the occurrence of the McGurk effect was negatively correlated with the activities in the precentral gyrus (Benoit et al., 2010). The analysis of effective connectivity in the present study indicated that the left IFG was positively related to the precentral gyri and the left STS in incongruent trials compared with congruent trials, while no increase in the connectivity of the left STS with the motor system was shown in incongruent trials. The occurrence of the McGurk effect may be controlled by the balance of activities between the left STS and the motor system (including the left IFG), with the STS promoting multisensory integration for inducing the McGurk effect, and the motor system recognizing audiovisual incompatibility and avoiding the McGurk effect.

Previous studies using TMS have demonstrated that the MEP amplitudes of relaxed lip muscles increase while viewing lip movements and listening to speech (Sundara et al., 2001; Watkins et al., 2003; Murakami et al., 2011). Double-pulse TMS over the articulatory M1 facilitates the identification of perceived syllables (D'Ausilio et al., 2009). Low-frequency repetitive TMS of the M1 lip area impairs the discrimination of lip-articulated syllables (Möttönen and Watkins, 2009). The corticocortical connectivity of the M1 lip area to the IFG increases while listening to speech conditions, but not in noise conditions (Watkins and Paus, 2004; Murakami et al., 2012). These findings represent convergent evidence that the M1 representations of the articulatory musculature contribute to the recognition of speech perception, which is in accordance with the motor theory of speech perception (Liberman and Mattingly, 1985). The present TMS experiments support this idea by showing that single-pulse TMS over the M1 lip areas reduced the occurrence of the McGurk effect in incongruent trials. TMS might enhance the role of the motor system, including the left IFG-precentral network, in modulating multisensory perception. The McGurk effect was modulated when TMS was applied over the M1 lip areas but not over the M1 foot area, indicating topographically specific control in the motor system. This result is consistent with previous TMS studies (Fadiga et al., 2002; Watkins et al., 2003; Murakami et al., 2011).

This is the first study demonstrating that TMS of the right M1 lip area modulates speech perception. The MEP of the lip muscle increased during both visual and auditory speech perceptions when TMS was applied over the left (but not the right) hemisphere (Watkins et al., 2003). Single-pulse TMS over the left (but

not the right) M1 was found to improve lexical decisions of action (Pulvermüller et al., 2005). This discrepancy might be due to differences in the task contents: the audiovisual stimuli in the present experiments consisted of syllables, and the precentral gyri activate bilaterally in phonetic perception tasks (Nishitani and Hari, 2002; Callan et al., 2004; Skipper et al., 2007), whereas Watkins et al. (2003) and Pulvermüller et al. (2005) used a context-dependent level of speech prose and lexical tasks, which are dominantly processed in the left hemisphere. Similar results were obtained when TMS was applied over the right M1 hand area, with the MEP amplitude increasing while observing simple movements of the left hand (Aziz-Zadeh et al., 2002).

We took advantage of the high temporal resolution of TMS to study the time course of network processing related to the emergence of the McGurk effect. Single-pulse TMS over the M1 lip area effectively reduced the McGurk effect from 100 ms before to 200 ms after the auditory onset. This finding supports the notion that modulatory roles of TMS in multisensory perception are temporally specific to the stimulus presentation. The temporal range was slightly longer than that in a previous TMS study reporting that TMS over the left STS disrupted the McGurk effect within the time window from 100 ms before to 100 ms after the auditory onset (Beauchamp et al., 2010). The effects of TMS on speech perception might be explained by evidence that the mirror neuron system forms a distributed neural network of the STS linking the motor system, including the IFG and M1 (Rizzolatti and Craighero, 2004). The STS is situated upstream of the motor system, and activation of the M1 follows STS activation via the left IFG by 150 ms during the observation of lip movements (Nishitani and Hari, 2002). This specific time window is consistent with previous behavioral findings that the McGurk effect appears within a 300 ms integration window of asynchronous auditory and visual stimulus onsets (van Wassenhove et al., 2007; Stevenson et al., 2012).

We took care to control for any nonspecific effect of TMS that might have influenced the McGurk effect. First, TMS over the M1 lip areas might have caused discomfort by contracting the temporal muscles, because the facial M1 areas are situated laterally from the M1 hand areas (Murakami et al., 2013). However, IS changed without changes in the RTs, suggesting genuine effects of TMS on multisensory perception. Second, it could be argued that click noise generated by TMS might have influenced the reduction of McGurk effect. We therefore used noise-isolating in-ear headphones to avoid interference from the click noise. Also, TMS was selectively applied over the target sites and its effects were observed 200 ms after the auditory onset, which would have minimized any contribution from nonspecific click sounds.

Another concern is related to using TMS to evaluate corticobulbar excitability. A previous study showed increases in the corticobulbar excitability during McGurk stimuli when recording MEP increases of facial muscles (Sato et al., 2010). We did not record MEPs in the present study because the aim of the TMS experiments was to clarify how the motor network modulates the McGurk effect. Future studies should investigate the relationship between the occurrence of the McGurk effect and corticobulbar excitability.

In conclusion, the findings of the present study suggest that the left IFG is part of the system that reduces illusion induced by multisensory incompatibility. The motor network consisting of the left IFG and precentral gyri contributes to the detection and resolution of multisensory incompatibility and plays a role in regulating speech perception.

References

Aziz-Zadeh L, Maeda F, Zaidel E, Mazziotta J, Iacoboni M (2002) Lateralization in motor facilitation during action observation: a TMS study. Exp Brain Res 144:127–131. CrossRef Medline

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

Beauchamp MS, Nath AR, 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. CrossRef Medline

Benoit MM, Raij T, Lin FH, Jääskeläinen IP, Stufflebeam S (2010) Primary and multisensory cortical activity is correlated with audiovisual percepts. Hum Brain Mapp 31:526–538. Medline

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

Callan DE, Jones JA, Callan AM, Akahane-Yamada R (2004) Phonetic perceptual identification by native- and second-language speakers differentially activates brain regions involved with acoustic phonetic processing and those involved with articulatory-auditory/orosensory internal models. Neuroimage 22:1182–1194. CrossRef Medline

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

Corballis MC (2010) Mirror neurons and the evolution of language. Brain Lang 112:25–35. CrossRef Medline

D'Ausilio A, Pulvermüller F, Salmas P, Bufalari I, Begliomini C, Fadiga L (2009) The motor somatotopy of speech perception. Curr Biol 19:381–385. CrossRef Medline

Fadiga L, Craighero L, Buccino G, Rizzolatti G (2002) Speech listening specifically modulates the excitability of tongue muscles: a TMS study. Eur J Neurosci 15:399–402. CrossRef Medline

Friston KJ, Buechel C, Fink GR, Morris J, Rolls E, Dolan RJ (1997) Psychophysiological and modulatory interactions in neuroimaging. Neuroimage 6:218–229. CrossRef Medline

Green AM, Angelaki DE (2010) Multisensory integration: resolving sensory ambiguities to build novel representations. Curr Opin Neurobiol 20:353–360. CrossRef Medline

Hickok G, Poeppel D (2007) The cortical organization of speech processing. Nat Rev Neurosci 8:393–402. CrossRef Medline

Hocking J, Price CJ (2008) The role of the posterior superior temporal sulcus in audiovisual processing. Cereb Cortex 18:2439–2449. CrossRef Medline

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

Matchin W, Groulx K, Hickok G (2014) Audiovisual speech integration does not rely on the motor system: evidence from articulatory suppression, the McGurk effect, and fMRI. J Cogn Neurosci 26:606–620. CrossRef Medline

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

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

Möttönen R, Watkins KE (2009) Motor representations of articulators contribute to categorical perception of speech sounds. J Neurosci 29:9819–9825. CrossRef Medline

Murakami T, Restle J, Ziemann U (2011) Observation-execution matching and action inhibition in human primary motor cortex during viewing of speech-related lip movements or listening to speech. Neuropsychologia 49:2045–2054. CrossRef Medline

Murakami T, Restle J, Ziemann U (2012) Effective connectivity hierarchically links temporoparietal and frontal areas of the auditory dorsal stream with the motor cortex lip area during speech perception. Brain Lang 122:135–141. CrossRef Medline

Murakami T, Ugawa Y, Ziemann U (2013) Utility of TMS to understand the neurobiology of speech. Front Psychol 4:446. CrossRef Medline

Murakami T, Kell CA, Restle J, Ugawa Y, Ziemann U (2015) Left dorsal speech stream components and their contribution to phonological processing. J Neurosci 35:1411–1422. CrossRef Medline

Nath AR, Beauchamp MS (2012) A neural basis for interindividual differ-

ences in the McGurk effect, a multisensory speech illusion. Neuroimage 59:781–787. CrossRef Medline

Nishitani N, Hari R (2002) Viewing lip forms: cortical dynamics. Neuron 36:1211–1220. CrossRef Medline

Novick JM, Trueswell JC, Thompson-Schill SL (2005) Cognitive control and parsing: reexamining the role of Broca's area in sentence comprehension. Cogn Affect Behav Neurosci 5:263–281. CrossRef Medline

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

Oldfield RC (1971) The assessment and analysis of handedness: the edinburgh inventory. Neuropsychologia 9:97–113. CrossRef Medline

Perini F, Cattaneo L, Carrasco M, Schwarzbach JV (2012) Occipital transcranial magnetic stimulation has an activity-dependent suppressive effect. J Neurosci 32:12361–12365. CrossRef Medline

Pulvermüller F, Hauk O, Nikulin VV, Ilmoniemi RJ (2005) Functional links between motor and language systems. Eur J Neurosci 21:793–797. CrossRef Medline

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

Rizzolatti G, Craighero L (2004) The mirror-neuron system. Annu Rev Neurosci 27:169–192. CrossRef Medline

Sack AT, Cohen Kadosh R, Schuhmann T, Moerel M, Walsh V, Goebel R (2009) Optimizing functional accuracy of TMS in cognitive studies: a comparison of methods. J Cogn Neurosci 21:207–221. CrossRef Medline

Sato M, Buccino G, Gentilucci M, Cattaneo L (2010) On the tip of the tongue: modulation of the primary motor cortex during audiovisual speech perception. Speech Commun 52:533–541. CrossRef

Schwarzkopf DS, Silvanto J, Rees G (2011) Stochastic resonance effects re-

veal the neural mechanisms of transcranial magnetic stimulation. J Neurosci 31:3143–3147. CrossRef Medline

Sekuler R, Sekuler AB, Lau R (1997) Sound alters visual motion perception. Nature 385:308. CrossRef Medline

Seltzer B, Pandya DN (1978) Afferent cortical connections and architectonics of the superior temporal sulcus and surrounding cortex in the rhesus monkey. Brain Res 149:1–24. CrossRef Medline

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

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

Stevenson RA, Zemtsov RK, Wallace MT (2012) Individual differences in the multisensory temporal binding window predict susceptibility to audiovisual illusions. J Exp Psychol Hum Percept Perform 38:1517–1529. CrossRef Medline

Sundara M, Namasivayam AK, Chen R (2001) Observation-execution matching system for speech: a magnetic stimulation study. Neuroreport 12:1341–1344. CrossRef Medline

van Wassenhove V, Grant KW, Poeppel D (2007) Temporal window of integration in auditory-visual speech perception. Neuropsychologia 45: 598–607. CrossRef Medline

Watkins K, Paus T (2004) Modulation of motor excitability during speech perception: the role of Broca's area. J Cogn Neurosci 16:978–987. CrossRef Medline

Watkins KE, Strafella AP, Paus T (2003) Seeing and hearing speech excites the motor system involved in speech production. Neuropsychologia 41: 989–994. CrossRef Medline