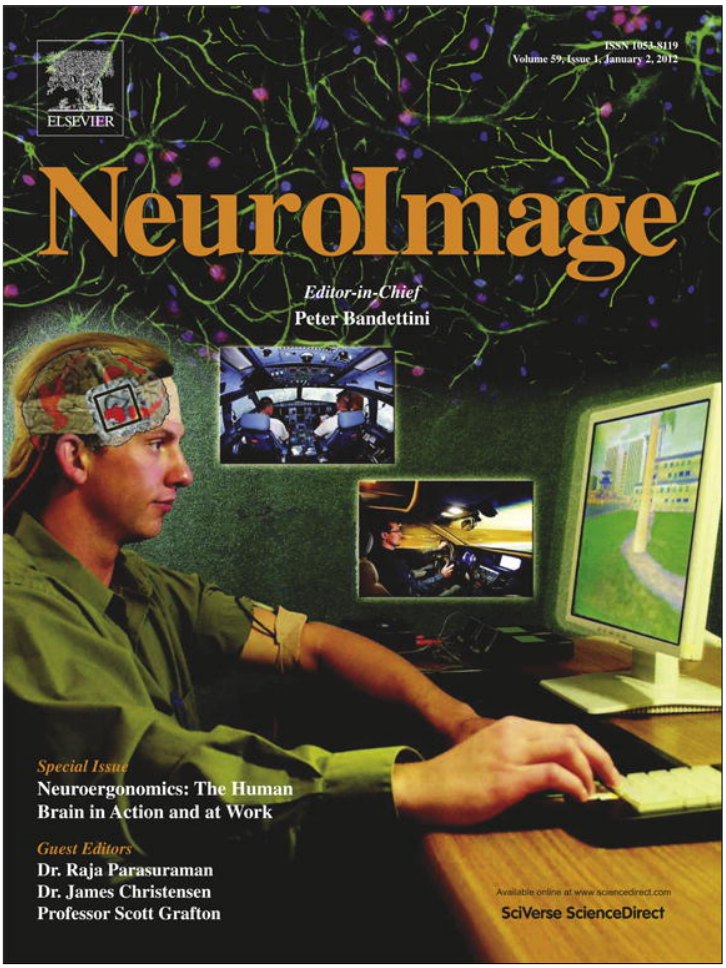
Provided for non-commercial research and education use.
Not for reproduction, distribution or commercial use.



This article appeared in a journal published by Elsevier. The attached copy is furnished to the author for internal non-commercial research and education use, including for instruction at the authors institution and sharing with colleagues.

Other uses, including reproduction and distribution, or selling or licensing copies, or posting to personal, institutional or third party websites are prohibited.

In most cases authors are permitted to post their version of the article (e.g. in Word or Tex form) to their personal website or institutional repository. Authors requiring further information regarding Elsevier's archiving and manuscript policies are encouraged to visit:

http://www.elsevier.com/copyright

Author's personal copy

NeuroImage 59 (2012) 781–787

Contents lists available at ScienceDirect

NeuroImage

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

A neural basis for interindividual differences in the McGurk effect, a multisensory speech illusion

Audrey R. Nath, Michael S. Beauchamp *

Department of Neurobiology and Anatomy, University of Texas Medical School at Houston, Houston TX, USA

ARTICLE INFO

Article history:
Received 8 April 2011
Revised 5 July 2011
Accepted 10 July 2011
Available online 20 July 2011

Keywords:
fMRI
McGurk
STS
Audiovisual integration
Speech

ABSTRACT

The McGurk effect is a compelling illusion in which humans perceive mismatched audiovisual speech as a completely different syllable. However, some normal individuals do not experience the illusion, reporting that the stimulus sounds the same with or without visual input. Converging evidence suggests that the left superior temporal sulcus (STS) is critical for audiovisual integration during speech perception. We used blood-oxygen level dependent functional magnetic resonance imaging (BOLD fMRI) to measure brain activity as McGurk perceivers and non-perceivers were presented with congruent audiovisual syllables, McGurk audiovisual syllables, and non-McGurk incongruent syllables. The inferior frontal gyrus showed an effect of stimulus condition (greater responses for incongruent stimuli) but not susceptibility group, while the left auditory cortex showed an effect of susceptibility group (greater response in susceptible individuals) but not stimulus condition. Only one brain region, the left STS, showed a significant effect of both susceptibility and stimulus condition. The amplitude of the response in the left STS was significantly correlated with the likelihood of perceiving the McGurk effect: a weak STS response meant that a subject was less likely to perceive the McGurk effect, while a strong response meant that a subject was more likely to perceive it. These results suggest that the left STS is a key locus for interindividual differences in speech perception.

© 2011 Elsevier Inc. All rights reserved.

Introduction

Humans combine the independent information available from the auditory modality (heard speech) and the visual modality (mouth movements) in order to aid in speech comprehension, our primary form of communication (Hickok and Poeppel, 2007; Poeppel et al., 2008; Samuel, 2011). A remarkable illusion known as the McGurk effect (McGurk and MacDonald, 1976) is a powerful demonstration of this integration: an auditory "ba" presented with the mouth movements of "ga" is perceived by the listener as a completely different syllable, "da" (the McGurk percept). However, the illusion is not experienced by all individuals, with population estimates of McGurk susceptibility ranging between 26% and 98% (Gentilucci and Cattaneo, 2005; McGurk and MacDonald, 1976). Why do only some people perceive the McGurk effect?

Converging evidence suggests that the STS is a critical brain area for multisensory integration of auditory and visual information about both speech and non-speech stimuli (Barraclough et al., 2005; Beauchamp et al., 2004; Callan et al., 2004; Calvert et al., 2000; Dahl et al., 2009; Macaluso et al., 2004; Miller and D'Esposito, 2005; Noesselt et al., 2007; Sekiyama et al., 2003; Stevenson and James, 2009; Werner and Noppeney, 2010). This suggests that the STS could be a neural locus for the McGurk effect: if the left STS successfully combines the incongruent auditory and visual syllables that comprise a McGurk stimulus, a McGurk percept is produced; if the left STS is not active, then the auditory and visual syllables are not combined and a McGurk percept is not produced. This idea received support from a recent study in which brain activity was disrupted with transcranial magnetic stimulation (TMS) (Beauchamp et al., 2010). When TMS was applied to the left STS of McGurk perceivers, the frequency of the McGurk percept was greatly reduced, rendering them more like non-perceivers. While this finding demonstrates that the left STS is necessary for McGurk perception in McGurk perceivers, it says nothing about the difference between perceivers and non-perceivers. Therefore, we formulated the hypothesis that a neural substrate for the difference between perceivers and non-perceivers would be differing activity levels in the left STS, and possibly other brain areas. Because increased activity in the STS is a neural signature for multisensory integration (Beauchamp et al., 2004; Van Atteveldt et al., 2004; Wright et al., 2003), our hypothesis predicted that during presentation of mismatched audiovisual speech, the STS should be strongly active in McGurk perceivers, reflecting their integration of the auditory and visual speech components, and only weakly active in McGurk non-perceivers, reflecting their lack of audiovisual integration. fMRI was used to measure brain activity in the STS and other critical areas for processing audiovisual speech: extrastriate visual cortex, auditory cortex, and inferior frontal gyrus (Broca's area) (Campbell, 2008; Scott and Johnsrude, 2003).

* Corresponding author at: 6431 Fannin St. Suite G.550 Houston, TX 77030, USA. Fax: +1 713 500 0623.
E-mail address: Michael.S.Beauchamp@uth.tmc.edu (M.S. Beauchamp).

1053-8119/$ – see front matter © 2011 Elsevier Inc. All rights reserved.
doi:10.1016/j.neuroimage.2011.07.024

Author's personal copy

A study of unisensory visual speech (Hall et al., 2005) found greater activity in auditory cortex in left superior temporal gyrus (STG) for subjects who were better at lip reading, supporting that idea that increased activity in temporal areas can reflect interindividual differences. In contrast, previous fMRI studies of the McGurk effect have found either no correlation between STS activity and McGurk susceptibility (Hasson et al., 2007; Jones and Callan, 2003) or a negative correlation (Benoit et al., 2010). A limitation of these studies is that they did not use independent functional localizers. In particular, the location in standard space of the posterior multisensory area in the STS varies greatly from subject to subject (Beauchamp et al., 2004) making it difficult to examine with standard group analysis techniques (Argall et al., 2006). Therefore, we used functional localizers to define the posterior multisensory area in the STS in each subject, and then examined the response of the STS to congruent and incongruent audiovisual speech in McGurk perceivers and non-perceivers.

Methods

Subjects and stimuli

14 healthy right-handed subjects (6 females, mean age 26.1) provided informed written consent under an experimental protocol approved by the Committee for the Protection of Human Subjects of the University of Texas Health Science Center at Houston.

A digital video was recorded of a female talker saying "ba, "ga", "da" and "ma" and edited with digital video editing software (iMovie, Apple Computer). The duration of the auditory syllables ranged from 0.4 to 0.5 seconds. The total length of each video clip ranged from 1.7 to 1.8 seconds in order to start and end each video in a neutral, mouth-closed position and to include all mouth movements from mouth opening to closing. The stimuli (Figs. 1A–C) consisted of three types of syllables: congruent (auditory and visual matching) syllables and two types of incongruent syllables (auditory and visual mismatch). Not all incongruent syllables produce a McGurk percept, defined as a percept not present in the original stimulus (McGurk and MacDonald, 1976). We created both McGurk (auditory "ba" + visual "ga" producing the McGurk percept of "da") and non-McGurk incongruent syllables (auditory "ga" + visual "ba" producing an auditory percept of "ga" or a combination percept such as "g-ba"). For a sample McGurk stimulus, please see the URL http://www.youtube.com/watch?v=WK3T7LWIkP8. View the video with eyes open and closed to experience the McGurk effect.

Behavioral pre-testing

Just prior to scanning, a behavioral pre-test was performed. Each subject was presented with 10 trials of McGurk syllables and 10 trials of non-McGurk incongruent syllables. Auditory stimuli were delivered through headphones at approximately 70 dB, and visual stimuli were presented on a computer screen. Subjects were instructed to watch the mouth movements and listen to the speaker. In order to assess perception, subjects were asked to repeat aloud the perceived syllable, with no constraints placed on potential responses: all responses were recorded exactly as spoken. This open-choice response has been shown to be a conservative measure of McGurk perception and is more informative with respect to possible intersubject differences in perception than a forced-choice response (Colin et al., 2005; Olson et al., 2002). Fused percepts such as "da" and "tha" were used as indicators of McGurk perception, while responses strictly corresponding to the visually-presented syllable ("ga") were not counted as fused McGurk percept. Responses corresponding to "ba" (the auditory component of the syllable) indicated that the effect was not perceived (McGurk and MacDonald, 1976).

fMRI functional localizer experiment

A functional localizer scan series was used to identify regions of interest (ROIs) important for speech perception in each subject; these ROIs were then applied to the independent data collected in the fMRI syllables experiment. The functional localizer scan series contained ten blocks (five unisensory auditory and five unisensory visual in random order) of duration 20 seconds with 10 seconds of fixation baseline between each block. Each block contained ten 2-second trials, one undegraded word per trial.

Word stimuli for the localizer were selected from two hundred single-syllable words from the MRC Psycholinguistic Database with Brown verbal frequency of 20 to 200, imageability rating greater than

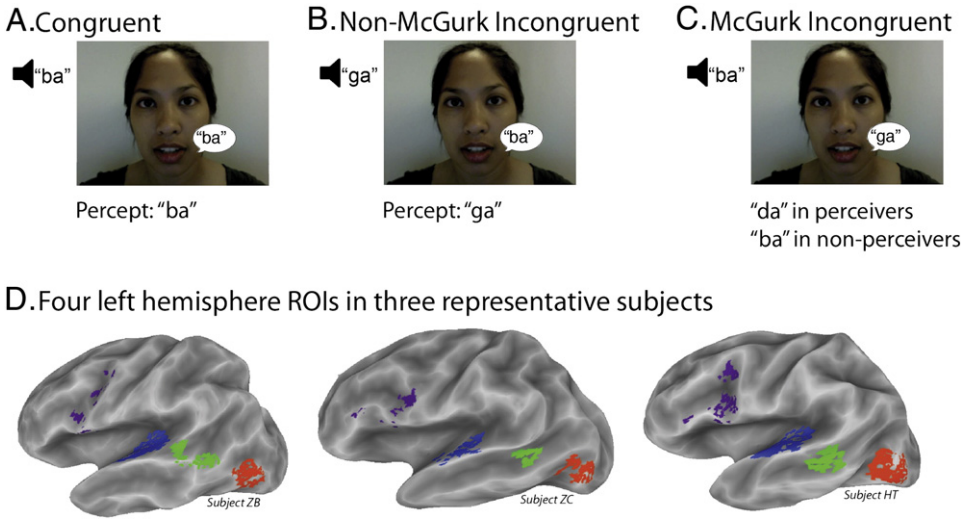


Fig. 1. Stimuli and regions of interest (ROIs). A. Congruent audiovisual syllable, consisting of matching auditory "ba" (depicted by speaker icon) and visual "ba" (single frame of video shown). Percept (shown below picture) is "ba." B. Non-McGurk incongruent syllable, consisting of auditory "ga" and visual "ba." This stimulus does not result in an illusory percept; the resulting percept is most often "ba." C. McGurk incongruent syllable, consisting of auditory "ba" and visual "ga." For McGurk perceivers, this results in the percept of an illusory "da." For non-perceivers, the percept is "ba." D. Four left hemisphere ROIs in three representative subjects. The STS ROI (green) contains voxels responsive to both auditory and visual words greater than baseline in the posterior STS. The auditory cortex ROI (blue) contains voxels responsive to auditory words greater than baseline within Heschl's gyrus. The visual cortex ROI (red) contains voxels responsive to visual words greater than baseline within extrastriate lateral occipitotemporal cortex. The IFG ROI (purple) contains voxels responsive to both auditory and visual words greater than baseline within the opercular region of the inferior frontal gyrus and the inferior portion of the precentral sulcus.

Author's personal copy

100, age of acquisition less than 7 years and Kucera-Francis written frequency greater than 80 (Wilson, 1988). The duration of the words ranged from 0.5 to 0.7 seconds. The total length of each video clip ranged from 1.1 to 1.8 seconds in order to start and end each video with the speaker in a neutral, mouth-closed position and to include all mouth movements from mouth opening to closing.

ROI creation

All analyses were conducted on ROIs defined individually in each subject. For each subject, active voxels within anatomically-defined regions were grouped to form four ROIs. Significant activation was defined solely using the block-design localizer experimental data to ensure independence. The STS ROI was defined using a conjunction analysis to find all voxels that responded to both auditory and visual words significantly greater than baseline in the anatomically-defined posterior STS ($q<0.05$ for each modality) (Beauchamp, 2005; Beauchamp et al., 2008). The auditory cortex ROI was defined using the contrast of auditory words *vs.* baseline to find active voxels within Heschl's gyrus. The visual cortex ROI was defined using the contrast of visual words *vs.* baseline to find active voxels within extrastriate lateral occipitotemporal cortex. The inferior frontal ROI was defined using a conjunction analysis to find all voxels that responded to both auditory and visual words greater than baseline within the automated FreeSurfer parcellation of the inferior frontal gyrus (pars opercularis) and inferior precentral sulcus, corresponding to portions of Brodmann areas 44 and 45 (Broca's area) (Fischl et al., 2004).

fMRI syllables experiment

Each subject underwent one fMRI scanning session. Each fMRI scan series lasted for 5 minutes, with 3 or 4 scan series collected from each subject. Within each scan series, single syllables were presented within 2-second trials using a rapid event-related design. Each trial contained a video of duration 1.7–1.8 seconds, with fixation crosshairs occupying the remainder of the trial. In five subjects, each scan series contained 55 McGurk trials, 55 non-McGurk incongruent trials, 10 target trials (audiovisual "ma") and 30 trials of fixation baseline. After scanning the first 5 subjects, we decided it would be worthwhile to incorporate an additional stimulus type, congruent speech. In the remaining nine subjects, each scan series contained 25 McGurk trials, 25 non-McGurk trials, 25 congruent "ga" trials (auditory + visual "ga"), 25 congruent "ba" trials, 10 target trials (audiovisual "ma") and 30 trials of fixation baseline. During fixation, the crosshairs were presented at the same position as the mouth during visual speech to minimize eye movements.

A catch trial design was used, similar to that used in many EEG or MEG experiments. On approximately 10% of trials, the target stimulus (audiovisual "ma") was presented. Subjects were required to respond to the target stimulus by pressing a button, but not to other stimuli. Target stimuli were analyzed separately from other stimuli. This ensures attention to the stimulus while preventing contamination of the brain response to non-target stimuli by motor planning or execution (Beauchamp et al., 2007). Subjects identified target syllables with high precision (98% accuracy) indicating attention to the stimuli.

MRI and fMRI analysis

At the beginning of each scanning session, two T1-weighted MP-RAGE anatomical MRI scans were collected at 3 Tesla using an 8-channel head gradient coil; the anatomical scans were aligned to each other and averaged to provide maximum gray–white contrast. Then, a cortical surface model was created with FreeSurfer (Dale et al., 1999; Fischl et al., 1999) and manipulating with SUMA (Argall et al., 2006). T2* weighted images for fMRI were collected using gradient-echo echo-planar imaging (TR = 2015 ms, TE = 30 ms, flip angle = 90°) with in-plane resolution of 2.75 × 2.75 mm. 33 3-mm axial slices were collected, resulting in whole-brain coverage in most subjects. Each functional scan series consisted of 153 brain volumes. The first three volumes, collected before equilibrium magnetization was reached, were discarded resulting in 150 usable volumes. MRI-compatible insert headphones (Sensimetrics, Inc., Malden, MA) were used to present high-fidelity auditory stimuli combined with external ear defender-type acoustic earmuffs to reduce scanner noise. Visual stimuli were projected onto a screen using an LCD projector and viewed through a mirror attached to the head coil. Behavioral responses were collected using a fiber-optic button response pad (Current Designs, Haverford, PA). MR-compatible eye tracking (Applied Science Laboratories, Bedford, MA) was used in all fMRI experiments to ensure alertness and visual fixation.

fMRI data analysis was carried out using Analysis of Functional NeuroImages software (AFNI) (Cox, 1996). Corrections for voxel-wise multiple comparisons were carried out using the false discovery rate procedure (Genovese et al., 2002) and reported as *q* values. Data w analyzed in each subject and then combined across subjects using a random-effects model. Functional data were aligned to the average anatomical dataset and motion-corrected for each voxel in each subject using a local Pearson correlation (Saad et al., 2009). All analyses were carried out in all voxels in each subject in the context of the generalized linear model using a maximum-likelihood approach using the AFNI function *3dDeconvolve*. Movement covariates and baseline drifts (as second-order polynomials, one per scan series) were modeled as regressors of no interest. All McGurk data was collected separately from the localizer using a rapid event related design. Deconvolution with tent functions was used to separately estimate the complete hemodynamic response function to each stimulus type in each voxel using nine tent functions that spanned the time between stimulus onset and 16 seconds after stimulus onset. No difference was observed between the two congruent conditions (congruent "ba" and "ga"), so these conditions were collapsed for further analysis.

We performed connectivity analyses to determine if changes in functional connectivity between language areas were correlated with McGurk susceptibility (Nath and Beauchamp, 2011). A structural equation model was constructed and tested for each subject. The model consisted of the four ROIs (auditory cortex, visual cortex, frontal cortex and STS) in the left hemisphere with bidirectional connections between auditory cortex and STS, between visual cortex and STS and frontal cortex and STS. The amplitude of the hemodynamic response was estimated for each individual McGurk stimulus and averaged within each ROI to produce a vector of 75–100 McGurk amplitudes. These amplitudes were used to calculate the correlation matrix and path coefficients in each subject using the AFNI functions *1ddot* and *1dsem*. The path coefficients obtained from each subject were correlated with each subject's McGurk susceptibility.

Results

Behavioral results

In the behavioral test, there was a high degree of intersubject variability in McGurk susceptibility (Fig. 2A). Three subjects never reported the McGurk percept (0% susceptibility) while two subjects always reported it (100% susceptibility). Subjects were classified into two groups: non-perceivers (seven subjects, susceptibility 0–49%) and perceivers (seven subjects, 50%–100%). In four subjects of the fourteen subjects from the scanned cohort, testing was performed both before and after scanning; susceptibility was similar in both testing sessions, with a mean difference in pre- and post-test scores of 5% ± 6.5% and no change in group assignment. During presentation of

Author's personal copy

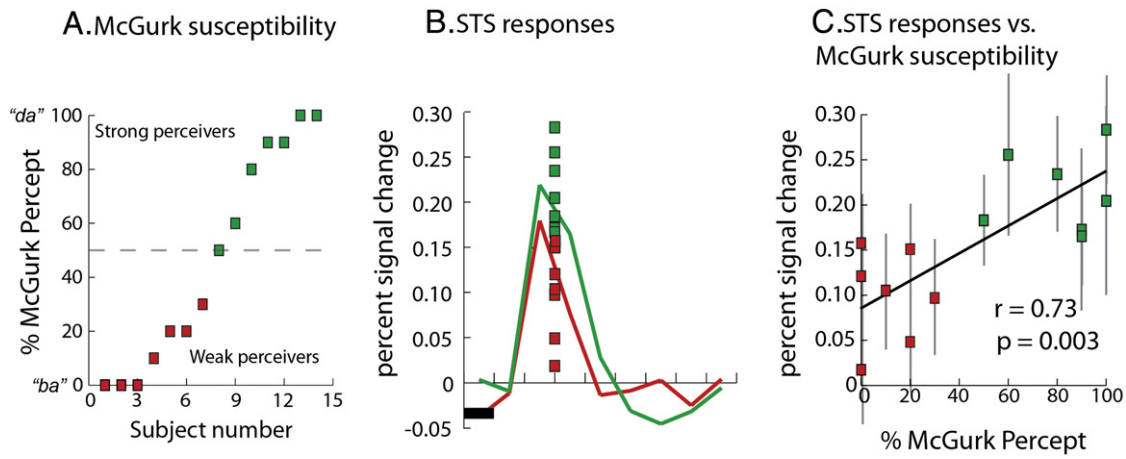


Fig. 2. McGurk susceptibility and STS responses during McGurk syllables. A. McGurk susceptibility for each of 14 subjects expressed as a percentage of responses corresponding to the McGurk percept during presentation of McGurk stimuli. Subjects were categorized as perceivers (green squares, ≥50%) and non-perceivers (red square, <50% McGurk susceptibility). B. Each square corresponds to the amplitude of response to McGurk stimuli in an individual subject's STS ROI, defined as the mean response between 4 seconds and 6 seconds after stimulus onset. The green and red tracings represent the average hemodynamic response curves across perceivers and non-perceivers, respectively. The black bar represents the time of stimulus onset (0 seconds); each tick mark is 2 seconds. C. The STS response to McGurk stimuli in each subjects are plotted against that subject's McGurk susceptibility. There was a significant positive correlation between each subject's STS response to McGurk syllables and their likelihood of experiencing the McGurk percept ($r = 0.73$, $p = 0.003$). Error bars represent standard deviation within each subject.

non-McGurk mismatched (incongruent) syllables, no subject reported a McGurk percept.

fMRI results

Functional localizers were collected in separate scan series, independent from the McGurk scan series, using a completely different stimulus set (unisensory words *vs.* audiovisual syllables) allowing statistical tests to be performed without bias. The word stimuli presented in the functional localizer scan series evoked robust activity in the four ROIs in each subject (Figs. 1D–F). While each subject showed a large region of STS that responded to both auditory-only and visual-only speech stimuli, after transformation into standard space the overlap across subjects was very small, reflecting both anatomical and functional variability in the location of the STS multisensory area. To quantify this variability, we measured the location of the STS ROI in each subject. The mean (±SD) center-of-mass was $x = -53.8 \pm 8.3$ mm, $y = -27.5 \pm 9.6$ mm, $z = 3.5 \pm 7.9$ mm. On average, each subject's STS center-of-mass was 13.7 mm from the mean center-of-mass, with some subjects more than 20 mm from the mean center-of-mass.

Using a rapid-event related design, we measured the brain response to presentation of congruent syllables and two types of incongruent syllables (McGurk and non-McGurk). A 2-way ANOVA was performed with BOLD response in the left STS as the dependent measure. The first factor was the McGurk susceptibility group determined from behavioral testing (perceivers vs. non-perceivers). The second factor was the stimulus condition (congruent vs. incongruent syllables). Because the division into groups was completely independent of the STS response, this analysis was unbiased. The ANOVA showed significant main effects of both susceptibility group ($F_{(1,17)} = 6.1$, $p = 0.03$) and stimulus condition ($F_{(1,17)} = 6.7$, $p = 0.02$) on the STS response. There was no interaction between susceptibility group and stimulus condition.

Effects of susceptibility

As shown in Fig. 2, McGurk perceivers had the highest STS response to incongruent syllables and non-perceivers had the lowest response ($0.22\% \pm 0.01\%$ *vs.* $0.12\% \pm 0.02\%$, $p = 0.001$). There was no between-group difference in the response to congruent syllables (perceivers *vs.* non-perceivers: 0.13% vs. 0.08%, $p = 0.22$).

Effects of stimulus

Across groups, the STS significantly preferred incongruent to congruent syllables (0.15% vs. 0.10%, $p = 0.002$). There was no significant difference between the responses to the two types of incongruent syllables (McGurk and non-McGurk).

Individual subject analyses

Next, we examined individual responses to the syllables (Fig. 2). The subject with the weakest STS response to mismatched speech (0.07%) had the smallest likelihood of experiencing a McGurk percept (0%); the subject with the strongest STS response (0.29%) had the highest likelihood (100%). Across all subjects, there was a significant positive correlation between each subject's STS response to incongruent syllables and their likelihood of experiencing the McGurk percept ($r = 0.73$, $p = 0.003$ for McGurk syllables and $r = 0.63$, $p = 0.02$ for incongruent syllables). There was no correlation between STS response and McGurk susceptibility for congruent syllables ($r = 0.42$, $p = 0.26$).

Additional analyses

A parallel analysis was conducted on four additional ROIs: left hemisphere inferior frontal gyrus (IFG), auditory cortex, extrastriate visual cortex, and right hemisphere STS. In the left IFG, the ANOVA showed a significant main effect of stimulus condition ($F_{(1,17)} = 18.1$, $p = 0.001$), but no effect of susceptibility group. In the auditory cortex, the ANOVA showed a significant main effect of susceptibility group ($F_{(1,17)} = 4.8$, $p = 0.04$), but no effect of stimulus condition. McGurk-susceptible individuals showed a trend towards greater auditory cortex response than McGurk-resistant individuals across stimulus types (0.26% vs. 0.18%, $p = 0.11$ for incongruent syllables; 0.24% vs. 0.14%, $p = 0.17$ for congruent syllables); there was no difference in auditory cortex response between incongruent and congruent syllables ($p = 0.34$ for non-perceivers, $p = 0.80$ for perceivers). Except for the left STS, no ROI showed a correlation between activity and McGurk susceptibility.

Author's personal copy

We considered whether changes in functional connectivity between the STS and frontal cortex, auditory cortex or extrastriate visual cortex could predict behavioral perception of McGurk stimuli. No correlation was observed between McGurk susceptibility and STS-frontal cortex connectivity ($r = -0.31$, $p = 0.28$), STS-auditory cortex connectivity ($r = 0.41$, $p = 0.15$) or STS-visual cortex connectivity ($r = 0.34$, $p = 0.23$) during perception of McGurk stimuli.

Discussion

We examined subjects who reported very different percepts when presented with the same physical McGurk stimulus, an incongruent pairing of auditory and visual syllables. Some subjects, the McGurk perceivers, almost always reported a McGurk percept, defined as a percept that corresponded to neither the auditory nor visual syllable. Other subjects, the non-perceivers, rarely reported this percept. To understand the neural substrates of these interindividual differences, we used fMRI to measure the brain response to congruent and incongruent audiovisual syllables. Only one brain region, the left STS, showed a significant effect of both susceptibility group and stimulus condition. Across subjects, the amplitude of the response in the left STS was significantly correlated with the likelihood of perceiving the McGurk effect: a weak STS response meant that a subject was less likely to perceive the McGurk effect, and a strong STS response meant that a subject was more likely to perceive it.

Evidence for a critical role of the left STS in McGurk perception is provided by a recent TMS study (Beauchamp et al., 2010). McGurk perceivers received single-pulse TMS to their left STS during presentation of McGurk stimuli. Subjects reported a dramatic reduction in perception of the McGurk effect, from 94% to 43%. Instead of experiencing the McGurk percept, subjects instead reported perceiving only the auditory syllable; control TMS did not have this effect. This demonstrates a causal relationship between activity in the STS and the audiovisual integration necessary for McGurk perception in McGurk perceivers. In effect, the TMS turned perceivers (with presumed high STS activity) into non-perceivers (with left STS activity disrupted by the TMS pulse).

These results can be integrated into psychological models of speech perception, such as the fuzzy logic model of speech perception (Oden and Massaro, 1978). This model assumes that the auditory and visual components of speech are evaluated individually before integration, and the combined information leads to a decision. In the context of the McGurk effect, articulatory mouth movements (visual cues) have greater influence in the setting of ambiguous auditory consonant information. The syllables “ba” and “da” can be difficult to distinguish with auditory cues alone, but are visually very different. Therefore, a visual “ga,” which is similar to a visual “da,” provides strong evidence that the auditory syllable was “da” and not “ba” (McGurk and MacDonald, 1976; Sekiyama, 1994), but only for McGurk perceivers whose STS integrates the auditory and visual information. McGurk non-perceivers assign little weight to the visual cue and experience only an auditory percept, corresponding to the lack of response in their left STS.

The significant interindividual differences in the McGurk effect that we observed are consistent with previous studies. 36% of our subjects had greater than 80% McGurk percept probability, consistent with the range of 26% to 98% observed in previous studies (Benoit et al., 2010; Gentilucci and Cattaneo, 2005; MacDonald et al., 2000; McGurk and MacDonald, 1976), and our mean McGurk percept probability of 46% (collapsed across subjects) is within the literature range of 32% to 94% (Baynes et al., 1994; Benoit et al., 2010; Bovo et al., 2009; Norrix et al., 2006; Olson et al., 2002; Sams et al., 1998).

There is a growing literature on the neural substrates for interindividual differences in language ability and perception (Aziz-Zadeh et al., 2010; Eisner et al., 2010; Ludman et al., 2000; Mei et al., 2008; Wong et al., 2007). Of particular relevance is an fMRI study by Hall and colleagues on speechreading or lipreading of unisensory visual speech (Hall et al., 2005). When good lipreaders were presented with unisensory visual speech, Hall and colleagues found more activity in left auditory cortex (superior temporal gyrus). In a striking parallel with the Hall et al. study, we observed greater activity in the auditory cortex in McGurk perceivers. While we know of no studies that have compared lipreading ability with McGurk susceptibility, these parallel brain imaging results suggest that they may be linked. Like many previous studies, the Hall et al. study examined only unisensory speech. While studies of auditory speech (Belin et al., 2004; Poeppel et al., 2004) and visual speech (Bernstein et al., in press) are important, an understanding of the McGurk effect requires presentation of both auditory and visual speech. Evidence suggests that the posterior STS is an important neural locus for association between auditory and visual speech with strong positive responses to both unisensory auditory and unisensory visual stimuli (Beauchamp et al., 2004; Van Atteveldt et al., 2004; Wright et al., 2003), as opposed to the negative responses to the non-preferred modality observed in unisensory areas (Laurienti et al., 2002).

In the present study, we tested the idea that the difference between McGurk perceivers and non-perceivers might be reflected in differences in the neural response in their left STS. Our finding supported this idea, with increased responses in the left STS for perceivers compared to non-perceivers. This finding has two key elements. First, the STS responses for both groups were similar for congruent audiovisual speech. Under normal conditions of congruent audiovisual speech, both groups activated the STS, allowing them to reap the benefit of the improved perceptual accuracy provided by multisensory integration. Second, the group difference was found for both McGurk syllables (for which the percept is different between perceivers and non-perceivers) and for non-McGurk incongruent syllables (for which the percept is similar for perceivers and non-perceivers). We speculate that McGurk perceivers have more liberal criteria for integrating auditory and visual speech information. Even if the auditory and visual information is mismatched, McGurk perceivers integrate it: this might provide an advantage under conditions of high levels of auditory or visual noise, at the cost of being misled by McGurk stimuli. Audiovisual integration is especially important under noisy conditions because multisensory integration is strongest for weak unisensory stimuli (Eisner et al., 2010; Grant and Seitz, 2000; Stein and Meredith, 1993). Sekiyama and Tohkura (1991) found heightened McGurk effect in the context of auditory noise. We have demonstrated that the connectivity of the STS with auditory and visual cortex is correlated with the degree of noise in each sensory modality for congruent audiovisual speech (Nath and Beauchamp, 2011). Therefore, a possible neural explanation for the Sekiyama and Tokhura result is that as the auditory noise increases, the connection of auditory cortex with STS decreases, and the connection of visual cortex with STS increases, heightening the influence of the visual stimulus and increasing the likelihood of a McGurk percept. In the present study, noisy McGurk stimuli were not presented, but this explanation could be tested in future experiments. An additional demonstration of the relevance of the McGurk effect for normal speech perception is that it can occur in the context of more naturalistic speech stimuli such as words and entire sentences (Sams et al., 1998).

Finally, we turn to the question of why three previous fMRI studies that differentiated subjects based on McGurk susceptibility (Benoit et al., 2010; Hasson et al., 2007; Jones and Callan, 2003) failed to find the positive correlation between STS activity and McGurk susceptibility observed in our experiment. A possible explanation for this disparity is that previous studies did not use independent functional localizers to identify the STS. The study by Jones and Callan (2003) used voxel-wise regression on fMRI data to search for voxels with a significant correlation between brain activity and McGurk susceptibility and did not report correlation in any STS voxels. However, because of intersubject variability, examining individual voxels in standard space may not compare functionally homologous regions between subjects. In the present study, the STS multisensory area was more than 2 cm from

Author's personal copy

the mean location in some subjects. The study by Hasson et al. (2007) used a repetition suppression paradigm to examine the fMRI response to different congruent syllables followed by a McGurk syllable. No differences between conditions were reported in the STS. However, the primary regions of interest (ROI) were defined anatomically. This presents a problem because the STS is the second largest sulcus in the human brain, after the Sylvian fissure. Because the STS multisensory area constitutes only a small portion of the entire STS, averaging across all voxels in the STS includes many voxels that have no response to speech stimuli, decreasing statistical power. This is illustrated in another study that also used repetition suppression of McGurk stimuli with an anatomical ROI consisting of the entire STS (Benoit et al., 2010). Our reanalysis of the Benoit et al. data found that their STS response to McGurk stimuli was not significantly different from zero ($p = 0.90$). In contrast, in our data, the STS response to McGurk stimuli (using an ROI from the independent functional localizer) was significantly greater than zero (mean response of 0.16%, $p = 0.000003$). In summary, previous studies did not use functional localizers to identify the location of the multisensory portion of STS in each individual subject. Group-wise analyses are insensitive to regions with high variability in standard space, such as the STS. Using functional localizers, we found a strong relationship between STS activity and McGurk susceptibility, supporting the idea that the STS is a critical brain locus for audiovisual integration in speech perception.

Acknowledgments

This research was supported by NSF642532, and NIHR01NS065395, TL1RR024147 and S10RR019186. We thank Vips Patel for assistance with MR data collection.

References

Argall, B.D., Saad, Z.S., Beauchamp, M.S., 2006. Simplified intersubject averaging on the cortical surface using SUMA. Hum. Brain Mapp. 27, 14–27.
Aziz-Zadeh, L., Sheng, T., Gheytanchi, A., 2010. Common premotor regions for the perception and production of prosody and correlations with empathy and prosodic ability. PLoS One 5, e8759.
Barraclough, N.E., Xiao, D., Baker, C.I., Oram, M.W., 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.
Baynes, K., Funnell, M.G., Fowler, C.A., 1994. Hemispheric contributions to the integration of visual and auditory information in speech perception. Percept. Psychophys. 55, 633–641.
Beauchamp, M.S., 2005. Statistical criteria in FMRI studies of multisensory integration. Neuroinformatics 3, 93–114.
Beauchamp, M.S., Lee, K.E., Argall, B.D., Martin, A., 2004. Integration of auditory and visual information about objects in superior temporal sulcus. Neuron 41, 809–823.
Beauchamp, M.S., Yasar, N.E., Kishan, N., Ro, T., 2007. Human MST but not MT responds to tactile stimulation. J. Neurosci. 27, 8261–8267.
Beauchamp, M.S., Yasar, N.E., Frye, R.E., Ro, T., 2008. Touch, sound and vision in human superior temporal sulcus. NeuroImage 41, 1011–1020.
Beauchamp, M.S., Nath, A.R., Pasalar, S., 2010. fMRI-guided transcranial magnetic stimulation reveals that the superior temporal sulcus is a cortical locus of the McGurk effect. J. Neurosci. 30, 2414–2417.
Belin, P., Fecteau, S., Bedard, C., 2004. Thinking the voice: neural correlates of voice perception. Trends Cogn. Sci. 8, 129–135.
Benoit, M.M., Raij, T., Lin, F.-H., Jaaskelainen, I.P., Stufflebeam, S., 2010. Primary and multisensory cortical activity is correlated with audiovisual percepts. Hum. Brain Mapp. 31, 526–538.
Bernstein, L.E., Jiang, J., Pantazis, D., Lu, Z.-L., Joshi, A., 2011. Visual phonetic processing localized using speech and nonspeech face gestures in video and point-light displays. Hum. Brain Mapp. doi:10.1002/hbm.21139.
Bovo, R., Ciorba, A., Prosser, S., Martini, A., 2009. The McGurk phenomenon in Italian listeners. Acta Otorhinolaryngol. Ital. 29, 203–208.
Callan, D.E., Jones, J.A., Munhall, K., Kroos, C., Callan, A.M., Vatikiotis-Bateson, E., 2004. Multisensory integration sites identified by perception of spatial wavelet filtered visual speech gesture information. J. Cogn. Neurosci. 16, 805–816.
Calvert, G.A., Campbell, R., Brammer, M.J., 2000. Evidence from functional magnetic resonance imaging of crossmodal binding in the human heteromodal cortex. Curr. Biol. 10, 649–657.
Campbell, R., 2008. The processing of audio-visual speech: empirical and neural bases. Philos. Trans. R. Soc. B 363, 1001–1010.
Colin, C., Radeau, M., Deltenre, P., 2005. Top-down and bottom-up modulation of audiovisual integration in speech. Eur. J. Cogn. Psychol. 17, 541–560.
Cox, R.W., 1996. AFNI: Software for analysis and visualization of functional magnetic resonance neuroimages. Comput. Biomed. Res. 29, 162–173.
Dahl, C.D., Logothetis, N.K., Kayser, C., 2009. Spatial organization of multisensory responses in temporal association cortex. J. Neurosci. 29, 11924–11932.
Dale, A.M., Fischl, B., Sereno, M.I., 1999. Cortical surface-based analysis. I. Segmentation and surface reconstruction. NeuroImage 9, 179–194.
Eisner, F., McGettigan, C., Faulkner, A., Rosen, S., Scott, S.K., 2010. Inferior frontal gyrus activation predicts individual differences in perceptual learning of cochlear-implant simulations. J. Neurosci. 30, 7179–7186.
Fischl, B., Sereno, M.I., Dale, A.M., 1999. Cortical surface-based analysis. II: Inflation, flattening, and a surface-based coordinate system. NeuroImage 9, 195–207.
Fischl, B., van der Kouwe, A., Destrieux, C., Halgren, E., Segonne, F., Salat, D.H., Busa, E., Seidman, L.J., Goldstein, J., Kennedy, D., Caviness, V., Makris, N., Rosen, B., Dale, A.M., 2004. Automatically parcellating the human cerebral cortex. Cereb. Cortex 14, 11–22.
Genovese, C.R., Lazar, N.A., Nichols, T., 2002. Thresholding of statistical maps in functional neuroimaging using the false discovery rate. NeuroImage 15, 870–878.
Gentilucci, M., Cattaneo, L., 2005. Automatic audiovisual integration in speech perception. Exp. Brain Res. 167, 66–75.
Grant, K.W., Seitz, P.F., 2000. The use of visible speech cues for improving auditory detection of spoken sentences. J. Acoust. Soc. Am. 108, 1197–1208.
Hall, D.A., Fussell, C., Summerfield, A.Q., 2005. Reading fluent speech from talking faces: typical brain networks and individual differences. J. Cogn. Neurosci. 17, 939–953.
Hasson, U., Skipper, J.I., Nusbaum, H.C., Small, S.L., 2007. Abstract coding of audiovisual speech: beyond sensory representation. Neuron 56, 1116–1126.
Hickok, G., Poeppel, D., 2007. The cortical organization of speech perception. Nat. Rev. Neurosci. 8, 393–402.
Jones, J.A., Callan, D.E., 2003. Brain activity during audiovisual speech perception: an fMRI study of the McGurk effect. Neuroreport 14, 1129–1133.
Laurienti, P.J., Burdette, J.H., Wallace, M.T., Yen, Y.F., Field, A.S., Stein, B.E., 2002. Deactivation of sensory-specific cortex by cross-modal stimuli. J. Cogn. Neurosci. 14, 420–429.
Ludman, C.N., Summerfield, A.Q., Hall, D., Elliott, M.R., Foster, J., Hykin, J.L., Bowtell, R., Morris, P.G., 2000. Lip-reading ability and patterns of cortical activation studied using fMRI. Br. J. Audiol. 34, 225–230.
Macaluso, E., George, N., Dolan, R., Spence, C., Driver, J., 2004. Spatial and temporal factors during processing of audiovisual speech: a PET study. NeuroImage 21, 725–732.
MacDonald, J.D., Andersen, S., Bachmann, T., 2000. Hearing by eye: how much spatial degradation can be tolerated? Perception 29, 1155–1168.
McGurk, H., MacDonald, J.W., 1976. Hearing lips and seeing voices. Nature 264, 746–748.
Mei, L., Chen, C., Xue, G., He, Q., Li, T., Xue, F., Yang, Q., Dong, Q., 2008. Neural predictors of auditory word learning. Neuroreport 19, 215–219.
Miller, L.M., D'Esposito, M., 2005. Perceptual fusion and stimulus coincidence in the cross-modal integration of speech. J. Neurosci. 25, 5884–5893.
Nath, A.R., Beauchamp, M.S., 2011. Dynamic changes in superior temporal sulcus connectivity during perception of noisy audiovisual speech. J. Neurosci. 31, 1704–1714.
Noesselt, T., Rieger, J.W., Schoenfeld, M.A., Kanowski, M., Hinrichs, H., Heinze, H.J., Driver, J., 2007. Audiovisual temporal correspondence modulates human multisensory superior temporal sulcus plus primary sensory cortices. J. Neurosci. 27, 11431–11441.
Norrix, L.W., Plante, E., Vance, R., 2006. Auditory–visual speech integration by adults with and without language-learning disabilities. J. Commun. Disord. 39, 22–36.
Oden, G.C., Massaro, D.W., 1978. Integration of featural information in speech perception. Psychol. Rev. 85, 172–191.
Olson, I.R., Gatenby, J.C., Gore, J.C., 2002. A comparison of bound and unbound audio-visual information processing in human cerebral cortex. Cogn. Brain Res. 14, 129–138.
Poeppel, D., Wharton, C., Fritz, J., Guillemin, A., San Jose, L., Thompson, J., Bavelier, D., Braun, A., 2004. FM sweeps, syllables and word stimuli differentially modulate left and right non-primary auditory areas. Neuropsychologia 42, 183–200.
Poeppel, D., Idsardi, W.J., van Wassenhove, V., 2008. Speech perception at the interface of neurobiology and linguistics. Philos. Trans. R. Soc. B 363, 1071–1086.
Saad, Z.S., Glen, D.R., Chen, G., Beauchamp, M.S., Desai, R., Cox, R.W., 2009. A new method for improving functional-to-structural MRI alignment using local Pearson correlation. NeuroImage 44, 839–848.
Sams, M., Manninen, P., Surakka, V., Helin, P., Katto, R., 1998. McGurk effect in Finnish syllables, isolated words, and words in sentences: effects of word meaning and sentence context. Speech Commun. 26, 75–87.
Samuel, A.G., 2011. Speech perception. Annu. Rev. Psychol. 62, 49–72.
Scott, S.K., Johnsrude, I.S., 2003. The neuroanatomical and functional organization of speech perception. Trends Neurosci. 26, 100–107.
Sekiyama, K., 1994. McGurk effect and incompatibility : a cross-language study on auditory–visual speech perception. Studies and essays. Behav. Sci. Philos. 14, 29–62.
Sekiyama, K., Tohkura, Y., 1991. McGurk effect in non-English listeners: few visual effects for Japanese subjects hearing Japanese syllables of high auditory intelligibility. J. Acoust. Soc. Am. 90 (4 Pt 1), 1797–1805.
Sekiyama, K., Kanno, I., Miura, S., Sugita, Y., 2003. Auditory–visual speech perception examined by fMRI and PET. Neurosci. Res. 47, 277–287.
Stein, B.E., Meredith, M.A., 1993. The Merging of the Senses. MIT Press.
Stevenson, R.A., James, T.W., 2009. Audiovisual integration in human superior temporal sulcus: inverse effectiveness and the neural processing of speech and object recognition. NeuroImage 44, 1210–1223.
Van Atteveldt, N., Formisano, E., Goebel, R., Blomert, L., 2004. Integration of letters and speech sounds in the human brain. Neuron 43, 271–282.

Author's personal copy

Werner, S., Noppeney, U., 2010. Superadditive responses in superior temporal sulcus predict audiovisual benefits in object categorization. Cereb. Cortex 20, 1829–1842.
Wilson, M., 1988. The MRC Psycholinguistic Database: Machine Readable Dictionary, Version 2. Behav. Res. Methods Instrum. Comput. 20, 6–11.
Wong, P.C., Perrachione, T.K., Parrish, T.B., 2007. Neural characteristics of successful and less successful speech and word learning in adults. Hum. Brain Mapp. 28, 995–1006.
Wright, T.M., Pelphrey, K.A., Allison, T., McKeown, M.J., McCarthy, G., 2003. Polysensory interactions along lateral temporal regions evoked by audiovisual speech. Cereb. Cortex 13, 1034–1043.