

Primary and Multisensory Cortical Activity is Correlated with Audiovisual Percepts

Margo McKenna Benoit,^{1,2*} Tommi Raij,¹
Fa-Hsuan Lin,^{1,3} Iiro P. Jääskeläinen,^{1,4} and Steven Stuffelbeam^{1,5}

¹*Athinoula A. Martinos Center for Biomedical Imaging, Massachusetts General Hospital, Charlestown, MA, USA*

²*Massachusetts Eye and Ear Infirmary, Harvard Medical School, Boston, MA, USA*

³*Institute of Biomedical Engineering, National Taiwan University, Taipei, Taiwan*

⁴*Laboratory of Computational Engineering, Helsinki University of Technology, Espoo, Finland*

⁵*Division of Health Sciences & Technology (HST), Massachusetts Institute of Technology/Harvard Medical School, Cambridge, MA, USA*

Abstract: Incongruent auditory and visual stimuli can elicit audiovisual illusions such as the McGurk effect where visual /ka/ and auditory /pa/ fuse into another percept such as/ta/. In the present study, human brain activity was measured with adaptation functional magnetic resonance imaging to investigate which brain areas support such audiovisual illusions. Subjects viewed trains of four movies beginning with three congruent /pa/ stimuli to induce adaptation. The fourth stimulus could be (i) another congruent /pa/, (ii) a congruent /ka/, (iii) an incongruent stimulus that evokes the McGurk effect in susceptible individuals (lips /ka/ voice /pa/), or (iv) the converse combination that does not cause the McGurk effect (lips /pa/ voice/ ka/). This paradigm was predicted to show increased release from adaptation (i.e. stronger brain activation) when the fourth movie and the related percept was increasingly different from the three previous movies. A stimulus change in either the auditory or the visual stimulus from /pa/ to /ka/ (iii, iv) produced within-modality and cross-modal responses in primary auditory and visual areas. A greater release from adaptation was observed for incongruent non-McGurk (iv) compared to incongruent McGurk (iii) trials. A network including the primary auditory and visual cortices, nonprimary auditory cortex, and several multisensory areas (superior temporal sulcus, intraparietal sulcus, insula, and pre-central cortex) showed a correlation between perceiving the McGurk effect and the fMRI signal, suggesting that these areas support the audiovisual illusion. *Hum Brain Mapp* 31:526–538, 2010. © 2009 Wiley-Liss, Inc.

Key words: fMRI; speech perception; auditory cortex; visual cortex; hearing; cochlear implant

Contract grant sponsor: Otolaryngology Research Training Grant (MEEI); Contract grant number: T32 DC000020-18; Contract grant sponsor: NIH; Contract grant number: R01 NS048279-01A1; Contract grant sponsor: The Finnish Cultural Foundation; Contract grant numbers: K08 MH067966, 5P41RR14075

Contract grant sponsors: The MIND Institute, The Academy of Finland, The Sigrid Jusélius Foundation.

*Correspondence to: Margo McKenna Benoit, Department of Otolaryngology, Massachusetts Eye and Ear Infirmary, 243 Charles Street, Boston, Massachusetts 02114, USA. E-mail: margo@nmr.mgh.harvard.edu

Received for publication 25 June 2007; Revised 18 July 2009; Accepted 20 July 2009

DOI: 10.1002/hbm.20884

Published online 24 September 2009 in Wiley InterScience (www.interscience.wiley.com).

INTRODUCTION

The merging of senses can modulate our overall perceptions. For example, multisensory perceptual changes have been demonstrated with simple stimuli such as visual light bursts paired with auditory tones [Shams et al., 2005; Watkins et al., 2006] and with more complex language paradigms where auditory speech discrimination is improved when the speaker's mouth is seen [Grant and Seitz, 2000; Sumbly and Pollack, 1954]. In prior neuroimaging studies, multisensory integration has been demonstrated in an extended network including the superior temporal sulcus (STS), intraparietal sulcus (IPS), inferior frontal gyrus (IFG), premotor cortex, and superior and inferior colliculi [Calvert and Lewis, 2004; Champoux et al., 2006; Musacchia et al., 2006; Wallace et al., 1993]. Integration effects are most noticeable when one sensory system is impaired or when a component of the audiovisual signal is degraded [Callan et al., 2003; Stein and Meredith, 1993; Sumbly and Pollack, 1954]. For example, in the hearing impaired (HI) population, visual contributions take on a larger role in speech perception [Grant and Seitz, 2000; Grant et al., 1998]. Advances in understanding the neural mechanisms of audiovisual integration have the potential to improve speech perception and production in hearing impaired and cochlear implant patients.

Integration effects are contingent upon the phonetic and temporal congruency of stimuli. Congruent audiovisual information can improve accuracy and shorten reaction time compared to unimodal stimuli [Dodd, 1977; Giard and Peronnet, 1999; Schroger and Widmann, 1998; Sumbly and Pollack, 1954]. In contrast, the incongruent pairing of auditory and visual stimuli slows reaction times and increases error rates in speech perception [Giard and Peronnet, 1999]. There is evidence that the incongruency of audiovisual stimuli can be reflected in many cortical areas, with greater activity for incongruent compared to congruent stimuli [Bushara et al., 2001; Jones and Callan, 2003; Miller and D'Esposito, 2005; Ojanen et al., 2005; Pekkola et al., 2006; Raji et al., 2000]. It is possible that incongruent audiovisual speech stimuli take more time and neural resources to process because most natural speech is congruent and temporally aligned.

Studies of congruent and incongruent audiovisual information processing in normal hearing subjects may clarify which additional areas are activated during processing of incongruent stimuli and how this relates to perception of speech and language.

During natural speech, visual cues typically precede auditory information by 80–100 ms [Munhall and Vatikiotis-Bateson, 2004]. Visual lip movements suggest a limited selection of possible auditory speech articulations [Sams et al., 1991], and when subsequent auditory information is incongruent with visual information, the perception is susceptible to mistakes in syllable identification. A well-known example of this type of perceptual mistake is the McGurk effect [McGurk and MacDonald, 1976]. The

McGurk effect is an audiovisual illusion whereby incongruent audiovisual combinations alter the phonetic perception of speech syllables. For example, the specific combination of visual /ka/ and auditory /pa/ results in the perception of a third syllable /ta/, distinct from either stimulus. The velar movement for the syllable /ka/ is ambiguous, since both /ka/ and /ta/ have a similar visual appearance. Combined with the auditory /pa/, subjects are more likely to "hear" /ta/ than /ka/. This perceptual event is variable across subjects. Some individuals perceive every McGurk stimulus as /ta/ while others experience the illusion on a minority of presentations. In addition, not every incongruent audiovisual pairing results in this type of fusion. Switching the syllables (visual /pa/ and auditory /ka/) results in an unfused awareness of both stimuli in nearly all subjects, sometimes described as /pka/. The McGurk effect has been extensively used in studies of multisensory speech processing as a measurable and behaviorally relevant index of audiovisual integration [Grant and Seitz, 1998; Grant et al., 1998; Jones and Callan, 2003; Jones and Jarick, 2006; Möttönen, 2002; Munhall et al., 1996; Sams et al., 1991]. This illusion enables researchers to maintain a constant stimulus (incongruent audiovisual information) while measuring a reproducible alteration in multisensory perception. We sought to investigate whether brain activity reflects the individual auditory and visual stimuli or the fused illusory percept.

Adaptation fMRI is a powerful technique that can detect subtle differences in neural processing. A population of neurons will show decreased activity in response to similar repeated stimuli and a subsequent "release from adaptation" when presented with a novel stimulus [for a review, see Grill-Spector et al., 2006]. In functional studies, adaptation could be responsible for the blood oxygen level dependent (BOLD) signal decreases, faster reaction times, and improved accuracy often noted in repetition priming experiments [Grill-Spector et al., 2006; Schacter and Buckner, 1998; Wiggs and Martin, 1998]. After adaptation, small subsequent stimulus changes cause an increase in the BOLD signal [Grill-Spector et al., 1999]. This is thought to reflect increased activity in small populations of neurons within a voxel that are more sensitive to the new stimulus as opposed to the previous stimulus [Grill-Spector and Malach, 2001; Grill-Spector et al., 1999; Grill-Spector et al., 2006]. This type of activity is not detectable in voxel-wise subtraction techniques because subpopulations of neurons within the same voxel may be active in two different conditions.

The current study is the first to investigate cortical activity in response to multisensory speech stimuli using the McGurk effect and adaptation fMRI. We presented trains of congruent stimuli to induce adaptation followed by a congruent or an incongruent stimulus. The incongruent stimuli were designed to induce a release from adaptation, reflected in an increased BOLD signal. We hypothesized that McGurk incongruent stimuli (sometimes perceived as a congruent syllable) should result in a lesser degree of

release from adaptation compared to the other incongruent condition. In addition, we asked whether the degree of perceptual fusion across subjects correlated with the change in BOLD signal in the activated brain areas.

MATERIALS AND METHODS

Subjects included 16 healthy, right-handed, English speaking adults (11 female, mean age 29.6 years, range 19–47 years). Each subject gave informed consent prior to participation. The study protocol and methods were approved by the institutional human studies review boards at the Massachusetts General Hospital and the Massachusetts Eye and Ear Infirmary.

Stimuli were short audiovisual movie clips. The visual component was the lower half of a female face mouthing the syllables /pa/ and /ka/ [Sams et al., 1991]. The stimuli subtended a visual angle of 11°, projected on a screen in the scanner through mirrors. The auditory component was the same person voicing the syllables /pa/ and /ka/ at a comfortable listening level for each subject (~75-dB SPL) presented through MRI-compatible headphones fitted around subjects' ears (MR CONFON, Magdeburg, Germany). The duration of all auditory stimuli was 172 ms and the duration of visual stimuli was 590 ms. The voice onset time (VOT) was 270 ms for audiovisual pairs including auditory /pa/ and 150 ms for audiovisual pairs including auditory /ka/. Initial lip movements began at ~50 ms. Mouth opening and voice onset occurred at 300 ms. Each movie clip lasted 850 ms.

Auditory and visual components were paired to make four different types of stimuli (Table I): (i) congruent pa (visual /pa/ auditory /pa/), (ii) congruent ka (visual /ka/ auditory /ka/), (iii) the incongruent stimulus likely to evoke the McGurk effect (IncM = Incongruent McGurk, visual /ka/ auditory /pa/), and (iv) the converse incongruent stimulus that does not produce the McGurk effect (IncN = Incongruent Non-McGurk, visual /pa/ auditory /ka/). Subjects were required to push one button if they thought the stimuli were “congruent” (voice and lips saying the same syllable) and a different button if they thought the stimuli were “incongruent” (voice and lips saying different syllables). The IncN combination was designed as a control that would elicit an unfused percept in all subjects. Only one aspect of the stimulus—auditory or visual—was changed to produce each type of incongruent movie. For example, a trial consisting of three congruent /pa/ movies followed by an IncN movie (visual /pa/ auditory /ka/) could be created by changing only the auditory component of the final movie from /pa/ to /ka/. This allowed independent investigation of small auditory and visual stimulus changes on activations in multisensory and primary sensory cortices.

Each trial consisted of a train of four audiovisual video clips. To induce adaptation, three congruent/ pa/ movies began each trial, followed by (i) a congruent /pa/, (ii) a

TABLE I. Audiovisual stimulus pairs and the proposed perceptual effects of congruent stimuli, McGurk incongruent stimuli (IncM) and non-McGurk incongruent stimuli (IncN)

Stimulus type	Visual syllable	Auditory syllable	Perception
Congruent pa	/pa/	/pa/	“pa”
Congruent ka	/ka/	/ka/	“ka”
Incongruent McGurk fusion (IncM)	/ka/	/pa/	“ta”
Incongruent Non-McGurk combination (IncN)	/pa/	/ka/	“pka”

novel congruent /ka/, or (iii–iv) an incongruent movie (IncM or IncN). The total time to present all four movies in each trial was 3,400 ms. During the 4,600 ms intertrial interval, a still face was visible with no sound, and subjects responded with a right hand button push whether they thought the fourth movie was congruent (index finger response) or incongruent (middle finger response), regardless of the syllable. This task ensured that subjects always paid the same level of attention to the fourth movie. Subjects were informed that they may hear and see a variety of syllables during the study. Pilot data revealed that any baseline stimulus other than a still face (cross hair, oval shape, scrambled face) resulted in a large response to the novelty of face presentation, an effect that is undesirable in the present paradigm to investigate subtle differences in syllable perception. In addition, the cross hair baseline condition resulted in a robust “off” response in areas thought to be involved in the resting state network [Gusnard et al., 2001]. Setting this resting state as an absolute baseline can lead to a consistent set of deactivations when an auditory or visual task is presented [Mazoyer et al., 2001; Shulman et al., 1997]. Any cognitive subtraction required by the baseline stimulus could produce misleading activations in voxels not involved in the speech perception task. Therefore we chose a still face for our “baseline” condition and used adaptation fMRI to reduce this confounding activity. There were six randomized presentations of each trial type per run, and runs lasted 4 min and 48 s. Each subject completed between four and eight runs.

Functional brain images were acquired in a 3 Tesla MRI scanner (Tim Trio, Siemens Medical Solutions, Erlangen, Germany) with a 12-channel head coil array. Using clustered volume acquisition [Edmister et al., 1999], the stimuli were presented between echo planar images (EPIs) without acoustic scanner interference. The whole-head acquisition occurred from 2.6 to 4.6 s after the last stimulus in a trial. A 2D gradient-echo EPI sequence was used to collect T2*-weighted image volumes with the following parameters: TR 2,000 ms with 8 s in between onsets of consecutive EPIs, TE 30 ms, flip angle 90°, slice thickness 5 mm with 1-mm gap, 25 slices, voxel size 3.1 × 3.1 × 5 mm³. The long TR was chosen to allow for measurement of the

hemodynamic response in the auditory cortex without confounding auditory stimulation caused by the response to scanner noise. The sparse sampling design also enabled us to present audiovisual stimuli with minimal contamination by scanner noise [Bandettini et al., 1998; Hall et al., 1999].

In the same session, two anatomical whole-brain T1-weighted MP-RAGE images were acquired for each subject. The cortical surfaces were reconstructed using FreeSurfer software (<https://surfer.nmr.mgh.harvard.edu/>) on the basis of the anatomical images, incorporating motion correction, intensity normalization, skull stripping, and hemisphere inflation. EPI images were motion-corrected and spatially smoothed using a kernel of full-width-half-maximum 8 mm, reconstructed using FS-FAST, and coregistered to and overlaid on the inflated surface of the anatomical scan [Dale et al., 1999; Fischl et al., 1999].

Behavioral data including reaction times were analyzed for each subject. The number of in-scanner trials was limited by time constraints due to the long TR and adaptation design. Separate behavioral trials outside the scanner were conducted to obtain more trials, and in all possible cases in-scanner and out-of-scanner data were combined in determining a subject's fusion rates. A response of "congruent" for congruent trials and "incongruent" for both the McGurk and non-McGurk incongruent trials was scored as correct in determining overall performance. When the IncM stimulus was keyed as congruent, the trial was considered "fused," indicating that the subject experienced the McGurk effect. We expected that most subjects would respond "incongruent" for the IncN stimuli and those responses to McGurk stimuli would be more varied. In addition, response reaction times were extracted from available out-of-scanner data. Reaction time data were analyzed separately for subjects who experienced the McGurk effect greater than 50% of the time and those who rarely experienced the McGurk effect. A student's *t* test was performed to look for differences between the two groups.

Individual BOLD data were analyzed using a general linear model. At the group level, a fixed effects analysis was carried out and the averaged data were overlaid using spherical intersubject averaging onto an average anatomical brain from the same subjects. We compared trials that ended in an incongruent IncM or IncN stimulus to trials with only congruent /pa/ stimuli, and performed a direct comparison of the two incongruent stimuli (IncN-IncM). The comparisons between incongruent and congruent trials were corrected for multiple comparisons using the false discovery rate (FDR) method at a level of $P = 0.01$ and the color maps are depicted with thresholds at $P < 0.05$ and $P < 0.01$. The subtraction of IncN-IncM is thresholded at $P < 0.05$ and $P < 0.01$, uncorrected. Cluster size for areas exceeding 20 voxels with a significance corresponding to $P < 0.01$ were collected along with significance of the *t* test, anatomical area, and Talairach coordinates.

To further investigate the relationship between BOLD signal and perception of the McGurk effect, an ROI analysis was performed on a group of regions that were acti-

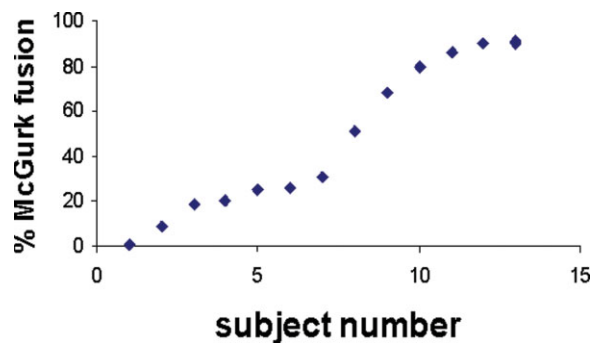


Figure 1.

Behavioral results for fusion of incongruent McGurk (IncM) stimuli into a congruent percept. The degree of fusion was ranked and the subject number reordered to show the continuum. Subjects at the lower end of the spectrum experienced the McGurk effect in as little as 1% of trials. In contrast, those at the higher end of the spectrum experienced the McGurk effect in up to 91% of trials. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

vated in the group statistical maps. Standard ROIs were delineated anatomically based on the FreeSurfer parcellation and then morphed from the standard space into each individual's brain via a spherical surface [Fischl et al., 2004]. This allowed ROIs to be standardized across subjects. The auditory cortex was subdivided into primary auditory cortex (Heschl's gyrus) and the nonprimary auditory areas in the superior temporal gyrus and supratemporal cortex (STG). For each subject, the BOLD percent signal change was extracted for each ROI in both the IncM and IncN conditions. A regression analysis was then performed comparing "percent fusion" (percent of McGurk incongruent trials keyed as congruent) and BOLD percent signal change. Plots were generated depicting the relationship with superimposed trendlines. The *P*-value of the *t*-statistic is reported.

RESULTS

Behavioral data were pooled from all runs for each subject. One subject had insufficient behavioral data and was excluded from further analysis. The remaining 15 subjects correctly categorized the IncN stimuli (iv) as incongruent in 87.8% of trials. The percent of McGurk stimuli reported as congruent for each subject is reported as "percent fusion" and plotted in Figure 1. The degree of fusion was ranked and the subject number reordered. The curve shows that subject responses fell along a continuum from a minimal fusion rate of 1% (most McGurk stimuli perceived as incongruent) to a maximum of 91% (most McGurk stimuli perceived as congruent). An analysis of reaction times revealed that those who responded to the McGurk stimuli as incongruent more than 50% of the time had a slightly longer processing time than those subjects

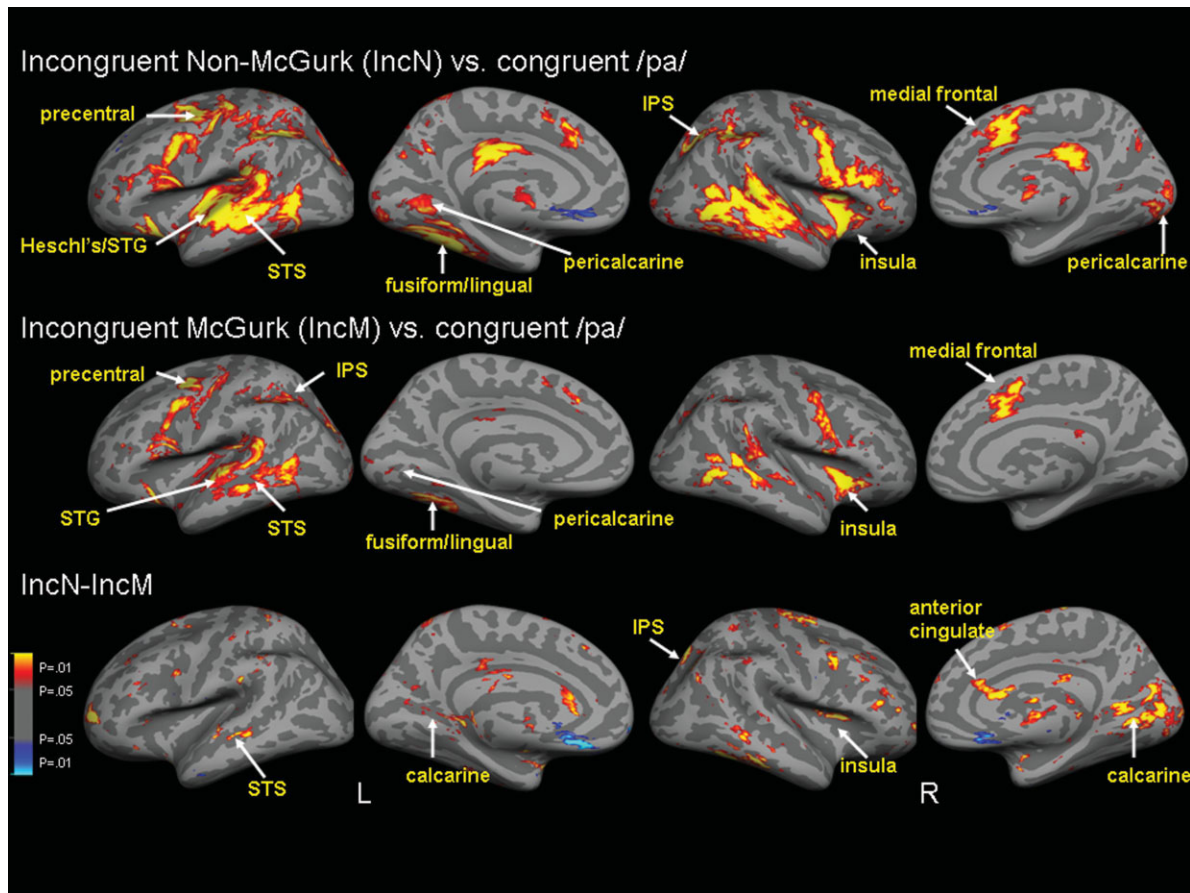


Figure 2.

Whole brain group activation maps on the lateral and medial surfaces of both hemispheres. All figures depict subtraction analyses. IncN = a subtraction of trials with four congruent /pa/ from trials with three congruent /pa/ followed by the Non-McGurk stimulus (visual /pa/ auditory /ka/). IncM = a subtraction of trials with four congruent /pa/ stimuli from trials with three congruent /pa/ followed by the McGurk stimulus (visual /ka/ audi-

tory /pa/). IncN-IncM is the subtraction of incongruent McGurk trials (visual /ka/ auditory /pa/) from non-McGurk trials (visual /pa/ auditory /ka/). All subtractions were done using a fixed effects model. The upper two rows were corrected for multiple comparisons using the false discovery rate method at 0.01. The IncN-IncM results are uncorrected. Thresholds were set according to the *P*-values shown in the color bar at the bottom left.

who responded that the McGurk stimuli were congruent more than 50% of the time (1,100 ms compared to 930 ms, $P < 0.05$).

Figure 2 top row shows the group analysis for a subtraction between incongruent non-McGurk trials (IncN) and congruent /pa/ trials. The IncN condition included an auditory change only (from congruent /pa/ to incongruent visual /pa/ auditory /ka/). There was a release from adaptation (increased BOLD signal) in the IncN condition in widespread areas including auditory (both Heschl's gyrus and STG), multisensory (STS, insula, precentral gyrus, intraparietal sulcus, anterior cingulate/medial superior frontal lobe, fusiform, and lingual gyri), and primary visual areas in both hemispheres. Talairach coordinates for significant clusters are listed in Table II.

Figure 2 middle row shows the group analysis for subtraction between incongruent McGurk trials (IncM) and congruent /pa/ trials. The IncM condition included a visual change only, from congruent /pa/ to incongruent visual /ka/ auditory /pa/. A similar network of auditory (STG) and multisensory areas (STS, precentral gyrus, right insula, IPS, medial superior frontal lobe, and left fusiform gyrus) showed release from adaptation; however, this appeared to be less robust than the IncN comparison. Much less BOLD activity was seen in primary visual cortex, despite the visual stimulus change.

To ensure that the type of change (auditory or visual) did not act as a confounding variable in our results, we also compared incongruent trials (IncN and IncM) to the congruent trials that consisted of three congruent /pa/

TABLE II. Talairach coordinates of areas significantly activated conditions IncN, IncM, and the subtraction of IncN-IncM in the left and right hemispheres

Condition	Anatomical description	Voxels	sig	X	Y	Z	
<i>Left Hemisphere</i>							
IncN	superior temporal sulcus and gyrus	12,223	11.75	-62.8	-21.0	0.0	
	precentral sulcus and gyrus	3,438	10.33	-28.3	-7.2	46.1	
	insula	1,137	9.05	-28.4	18.8	-10.8	
	intraparietal sulcus	3,639	7.73	-48.7	-40.7	42.5	
	main cingulate	463	5.74	-4.2	17.3	40.0	
	medial occipitotemporal gyrus (pericalcarine)	456	4.86	-9.4	-77.7	-5.3	
	occipital pole	196	4.53	-13.6	-96.0	-7.8	
	IncM	precentral sulcus	962	8.42	-26.8	-6.8	45.3
		fusiform gyrus	929	7.18	-39.6	-74.3	-7.3
		superior temporal sulcus/middle temporal gyrus	2,070	6.96	-58.5	-34.7	-1.5
insula		496	6.79	-28.2	23.8	-3.7	
superior temporal gyrus		2,747	6.56	-64.6	-17.0	0.5	
intraparietal sulcus		1,705	6.39	-43.4	-43.9	40.9	
precentral sulcus and gyrus		2,796	6.26	-39.7	4.7	36.8	
IncN-IncM		frontomarginal sulcus	174	4.07	-30.9	51.7	2.0
		Insula	124	2.95	-32.2	4.6	-16.2
		supramarginal gyrus	37	2.77	-62.8	-32.0	31.8
	calcarine sulcus	72	2.63	-18.9	-50.9	2.4	
<i>Right Hemisphere</i>							
IncN	superior temporal sulcus and gyrus	9,320	11.61	50.4	-33.3	0.3	
	insula	2,798	9.97	34.9	23.0	2.4	
	main cingulate	1,851	9.74	9.0	16.3	32.5	
	precentral sulcus and gyrus	4,419	8.33	40.4	4.8	31.3	
	intraparietal sulcus	2,730	6.86	43.4	-45.8	39.7	
	lingual gyrus	302	5.75	11.1	-87.5	-4.7	
	occipital pole	315	4.65	13.6	-92.1	3.6	
	IncM	insula	1,457	10.20	35.8	19.4	2.7
superior temporal sulcus		1,396	9.63	53.6	-42.0	7.9	
main cingulate		1,438	8.37	8.3	10.0	46.6	
precentral sulcus and gyrus		2,317	6.31	40.3	17.8	2.7	
superior temporal gyrus		1,356	6.30	66.4	-40.6	22.9	
occipital pole		185	6.08	24.4	-88.1	-5.2	
intraparietal sulcus		318	4.99	30.9	-63.6	31.0	
IncN-IncM		lateral occipito-temporal sulcus/fusiform	168	4.5	47.50	-53.0	-7.5
	superior parietal gyrus	174	3.29	20.5	-70.5	48.1	
	lingual gyrus	98	3.20	2.7	-67.7	6.5	
	calcarine sulcus	190	3.18	23.7	-55.9	9.2	
	precentral sulcus	35	2.78	34.4	2.3	46.7	
	posterior calcarine sulcus	109	2.77	7.8	-85.7	5.1	
	insula	101	2.60	33.4	20.4	8.5	
	occipital pole	22	2.38	12.2	-102.3	11.2	

IncN = incongruent non-McGurk condition (visual /pa/ auditory /ka/); IncM = incongruent McGurk condition (visual /ka/ auditory /pa/); IPS = intraparietal sulcus, MTG = middle temporal gyrus, STG = superior temporal gyrus, STS = superior temporal sulcus; sig = significance of the f-statistic.

syllables followed by a congruent/ka/(CpCk). In this case, the congruent condition resulted in a change in both auditory and visual stimuli. This allowed us to compare conditions in which there was an auditory change with an unfused perception (IncN) to trials in which there was an auditory change with a fused perception (CpCk). There was a similar pattern of activation in multisensory areas such as the insula, STS, and anterior cingulate, and in primary visual cortex for this subtraction (not shown). Simi-

larly a comparison of a visual change with variable fused perception (IncM) to a visual change with fused perception (CpCk) resulted in activity in multisensory and primary visual areas. These results suggest that the activations shown in Figure 2 are not due to the change in auditory or visual stimulus alone but are related to perceptual differences.

To further investigate the difference between McGurk and non-McGurk conditions, Figure 2 bottom row shows

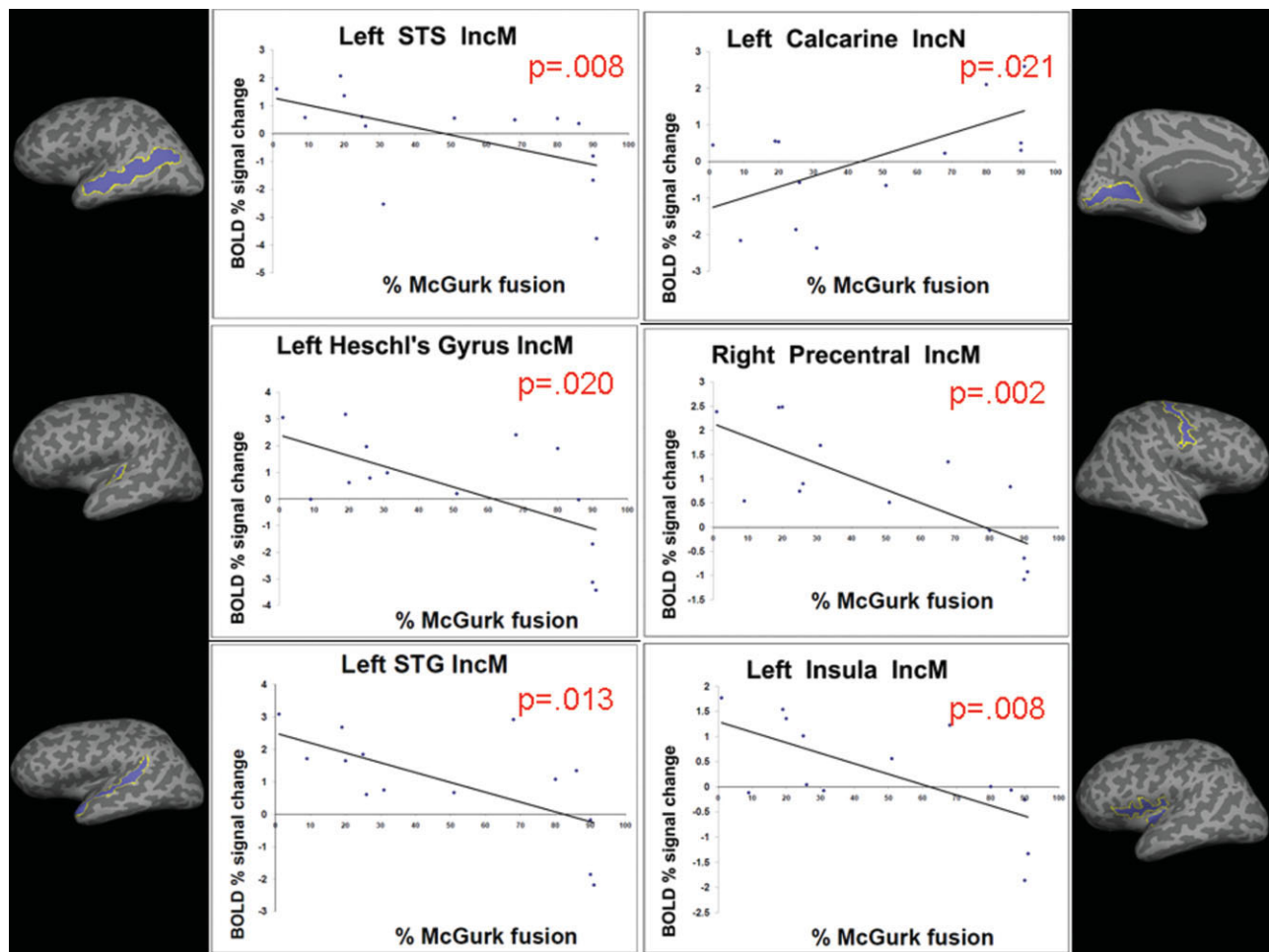


Figure 3.

ROI correlation analysis. Plots with trendlines are shown for each ROI where the degree of McGurk fusion correlated with the BOLD percent signal change. In the calcarine sulcus, there was a positive correlation between McGurk fusion and BOLD signal change for the IncN condition; for all other areas there was a negative correlation between McGurk fusion and BOLD signal change during the McGurk condition (IncM). Thumbnails of the ROIs are shown adjacent to the regression plots.

IncM trials (three congruent /pa/ followed by the McGurk visual /ka/ auditory /pa/) subtracted from IncN trials (three congruent /pa/ followed by the non-McGurk visual /pa/ auditory /ka/). This comparison revealed significant differences in the left STS, right insula, right IPS, bilateral anterior cingulate, and bilateral primary visual cortex.

For both auditory and visual changes, a release from adaptation was also seen in the sensory cortex of the opposite modality. For the IncN comparison with a change in only the auditory stimulus, there was a release from adaptation in pericalcarine areas bilaterally (Fig. 2, top row). For the IncM condition with only a visual change, there was a release from adaptation in the STG bilaterally (Fig. 2, middle row).

Figure 3 shows the results of the ROI correlation analysis. The thumbnails at the sides illustrate the ROIs plotted on an individual brain. There were significant correlations between the subject's behavioral data and the BOLD response to the McGurk stimulus (IncM vs. congruent /pa/) in primary and multisensory cortices. In the left primary auditory cortex (Heschl's gyrus), left STG (auditory association cortex), left STS, left IPS (not shown), left insula, and right precentral gyrus there was an inverse relationship between percent fusion and percent BOLD signal change in the IncM condition. In the left calcarine sulcus there was a positive correlation between the degree of McGurk fusion and BOLD response in the IncN condition.

DISCUSSION

In the present study we investigated neural correlates of phonemic perception using the McGurk audiovisual illusion in an adaptation fMRI design. Our experimental question was fundamentally different than asking where in the brain multisensory integration takes place. Prior research on the McGurk effect has been directed either at characterizing the behavioral phenomenon or at understanding the neural correlates of multisensory integration. Our study is unique in showing a correlation between perception of the McGurk effect and a measurable BOLD response, highlighting a network of cortical regions in the brain where multisensory speech perception takes place, independent from changes in the physical stimuli themselves.

Localizing the areas that correlate with behavioral perception lends further insight into the process of audiovisual speech comprehension. By adulthood, normal hearing individuals have acquired years of experience processing multisensory speech, and most develop a tendency to integrate simultaneous auditory and visual information even when there are discrepancies between the two modalities [Fowler, 2004]. When simultaneous auditory and visual inputs are perceived as incongruent, additional neural processing may be required. This is particularly relevant for HI patients, where altered auditory signals (from hearing aids or cochlear implants) may be perceived to be incongruent with visual information, requiring more time and neural resources for speech comprehension.

Studies that alter the congruency of multisensory conditions have consistently found greater activation for incongruent compared to congruent stimuli [Bushara et al., 2001; Jones and Callan, 2003; Miller and D'Esposito, 2005; Ojanen et al., 2005; Pekkola et al., 2006; Raji et al., 2000]. Our study extends these observations by demonstrating an incongruency effect for perceptually unfused (IncN) compared to perceptually fused (IncM) stimuli. The reaction times support this concept as well, with increased processing time required for subjects to respond to stimuli which were perceived as incongruent compared to the time required to respond to stimuli which were perceived as congruent syllables. In the present study, the whole brain analysis revealed that there was a greater BOLD percent signal change for incongruent stimuli than for congruent stimuli, and that McGurk stimuli (incongruent but sometimes perceived as congruent) fell in between. Incongruent audiovisual stimuli (IncN and IncM) induced a release from adaptation in both primary sensory and multisensory cortical areas, and the degree of adaptation release was reflected in the percept.

Prior anatomical and functional imaging studies have demonstrated that the STS is important for multisensory processing [Beauchamp et al., 2004; Benevento et al., 1977; Falchier et al., 2002; Kaas and Collins, 2004; Kaas and Hackett, 2000; Miller and D'Esposito, 2005; Watkins et al., 2006], particularly for language related stimuli [Callan et al., 2003; Calvert et al., 2000; Campbell et al., 2001; Levä-

nen et al., 2001; Raji et al., 2000; van Atteveldt et al., 2004]. Our results showing greater STS activity for incongruent stimuli are, however, contrary to some fMRI reports [Calvert et al., 1999, 2000] in which the STS showed increased activity for congruent stimuli and decreased activity for incongruent stimuli. These studies may not be directly comparable to our findings because they compare the multisensory response to the predicted response from the addition of independent unimodal auditory and visual stimuli. In contrast, our study investigated the effect of incongruent stimuli on neurons already adapted to congruent audiovisual stimuli, probing for a response to subtle syllable changes.

Additional regions that showed greater activity for incongruent compared to congruent stimuli were the STG (auditory association area), IPS, fusiform gyrus, insula, and precentral gyrus—all areas that have been previously implicated in multisensory processing [Benevento et al., 1977; Bushara et al., 2001; Calvert, 2001; Calvert and Lewis, 2004; Kaas and Collins, 2004; Kaas and Hackett, 2000; Miller and D'Esposito, 2005]. Incongruency in phonetically conflicting compared to matching vowels has been linked to activity in Broca's area and the left premotor cortex [Ojanen et al., 2005]. Perceptual incongruency due to temporal offset of audiovisual speech stimuli has also been shown to activate an extensive multisensory network compared with synchronous stimuli [Jones and Callan, 2003; Miller and D'Esposito, 2005]. In an event-related fMRI study on the perception of audiovisual temporal synchrony, Miller and D'Esposito [2005] reported distinct patterns of neural activity in the primary auditory cortex, STS, IPS, and inferior frontal gyrus. In our study, adaptation revealed a similar multisensory network with differential responses based on the phonetic perception of congruency.

In addition to multisensory effects, subjects showed different patterns of activity in primary sensory cortices for incongruent audiovisual stimuli. These effects differed based on percepts. The BOLD response to a change from congruent /pa/ to the IncM stimulus was small, reflecting an extension of the adaptation effect to similar nonidentical stimuli [Sawamura et al., 2006] or a more general adaptation to the congruency of audiovisual stimuli. When presented with a stimulus consistently perceived as incongruent (IncN), there was a release from adaptation in both primary auditory and visual areas as well as multisensory areas. This effect could be mediated by top-down differential feedback loops from multisensory cortex to primary sensory areas.

Prior anatomical and functional research has identified a network of areas involved in audiovisual processing with projections that extend from primary sensory areas to association areas and vice versa, along with direct connections between primary auditory and visual cortex [Benevento et al., 1977; Kaas and Hackett, 2000; Falchier et al., 2002]. Our results support previous work regarding visual influence on auditory processing during speech

discrimination [Bertelson et al., 2003; Calvert and Campbell, 2003; Calvert et al., 1997; Fingelkurtz et al., 2003; Finney et al., 2001; Hayes et al., 2003; Jones and Jarick, 2006; Levänen et al., 2001; MacSweeney et al., 2002; Massaro, 2004; Möttönen, 2002; Pekkola et al., 2005; Petitto et al., 2000]. However, natural speech discrimination is inherently multimodal, and perceptual differences likely reflect multisensory integration rather than merely visual interference on the auditory processing stream [Grant et al., 1998]. Interestingly, our results show a clear response in the calcarine sulcus during changes in the auditory speech component despite an identical visual stimulus (IncN) (Fig. 2 top row). Similar effects of auditory input on early visual processing have been shown in studies using basic audiovisual stimuli such as bleeps and light flashes [Watkins et al., 2006] or colored circles and tones [Molholm et al., 2002]. In addition, previous work in blind subjects [Roder et al., 2002] demonstrated activity in V1 when subjects listened to spoken sentences. It is likely that the same anatomical projections from multisensory areas to V1 are present in normal individuals [Falchier et al., 2002]; however, this influence on the earliest visual processing centers has not previously been shown in healthy adults for language related stimuli. These effects may have been subtracted out in previous voxel-wise fMRI comparisons because V1 is strongly activated by still faces (often serving as the control condition), whereas other extrastriate visual areas such as V5 and MT are not [Callan et al., 2003; Calvert and Campbell, 2003]. By using adaptation to moving faces and observing neural responses to perceptual phoneme changes, we were able to see these effects extending beyond V5/MT and into earlier visual areas.

Once this network of primary sensory and multisensory areas involved in the perception of incongruent McGurk stimuli was identified via whole brain analysis, we probed regions of interest more intensely for an association between behavioral fusion and BOLD activity. The behavioral curve of responses to the McGurk illusion (Fig. 1) shows a continuum of perceptual change from subjects who experience the McGurk effect greater than 90% of the time to those who rarely to never experience the fusion of the two audiovisual syllables into a coherent third syllable. This graded effect is similar to psychometric curves in other areas of neuroscience, particularly those involved with perception [Raizada and Poldrack, 2007; Xu, 2008]. The degree of variation among subjects allowed for further analysis of the relationship between perception and brain function by correlating the strength of the behavioral effect with the strength of BOLD activity in regions of interest, as shown in Figure 3. The correlation analysis revealed an inverse relationship between behavioral fusion and BOLD percent signal change in the McGurk condition in multisensory areas such as the left STS, insula, IPS, and the right precentral gyrus. The primary auditory cortex (Heschl's gyrus) and nonprimary auditory areas (STG) also demonstrated an inverse relationship. This result is consistent with earlier MEG work by Sams et al. showing

activity in primary auditory cortex during the McGurk effect [Sams et al., 1991]. Activity in these brain regions appears to be modified by perception of audiovisual congruency. Subjects who experienced McGurk fusion more often had less release from adaptation while those who perceived most McGurk stimuli as incongruent had a robust BOLD response. The inverse relationship supports the above findings that audiovisual stimuli perceived as incongruent induce a greater release from adaptation. This is true when only the perception of congruency changes despite a constant incongruent audiovisual stimulus. This finding implies that the increase in BOLD signal is due to additional neural resources required to clarify a discrepant perception, rather than a simple response to a change in components of the stimuli.

In the left calcarine sulcus, there was a positive correlation between McGurk perception and BOLD signal change for the IncN condition. Subjects who perceived the McGurk stimuli as congruent had more BOLD activity during the IncN condition, while those that perceived the McGurk stimuli as incongruent showed a smaller BOLD signal change for this condition. This result could be due to additional release from adaptation for subjects who perceive the McGurk stimuli as congruent and are then exposed to the IncN incongruent stimuli. Alternatively, this result could represent inhibitory top-down effects on the primary visual cortex following presentation of perceptually incongruent stimuli, decreasing reliance on the perceived incongruent or inconsistent visual modality. Alternative imaging methods with better temporal resolution may be required for further investigation of this effect.

Recent work by Raizada and Poldrack [2007] also showed alterations in BOLD response that were linked to behavioral perception. The authors measured selective amplification in the brain during perception of categorical phonetic perception. Using a behaviorally weighted general lineal model, they found correlated activity in six regions including the left inferior supramarginal gyrus, right cerebellum, anterior cingulate, left IPS, left middle frontal cortex, and right prefrontal cortex. These areas proposed to be important for phonetic perception overlapped with our results in the anterior cingulate, IPS, and middle frontal cortex. Our results also included a greater network of speech processing areas, perhaps because the task was directly related to understanding and judging speech syllables.

Jones and Callan [2003] have also investigated the effects of audiovisual perception on brain functioning. The authors studied the effects of altering congruency (using congruent and incongruent McGurk stimuli) and temporal synchrony. The McGurk effect was strongest when the audiovisual stimuli were temporally synchronous. Similar to our current findings, a comparison of the synchronous conditions showed greater activity in the right supramarginal gyrus and left inferior parietal lobule for incongruent compared to congruent stimuli. They did not find a correlation between perceptual performance and activity in the STS or auditory cortex. The authors did not report the degree of McGurk

fusion across subjects, and it is possible that their subject group was homogenous in the degree to which they were susceptible to the McGurk effect. In our study subjects' response to the McGurk effect varied across a continuum from 1% to 91% fusion. Because there is wide variation in the degree to which subjects perceive the McGurk effect, a correlation was apparent in our experiment.

Perceptual fusion depends on the syllable pair, the timing offset, and on the individual subject [Champoux et al., 2006; Massaro, 2004; McGurk and MacDonald, 1976]. The average McGurk fusion rate in this study (49%) was comparable with earlier published fusion rates ranging from 40% to near 100% [Hayes et al., 2003; McGurk and MacDonald, 1976; Möttönen, 2002; Munhall et al., 1996; Olson et al., 2002; Saldana and Rosenblum, 1993; Sams et al., 1991]. While our stimulus model was a native Finnish speaker and subjects were native English speakers, it is unlikely that differences in fusion between subject groups were due to cultural differences in syllable production [see Munhall and Vatikiotis-Bateson, 2004 for a review of the spatial constraints of speech information], and indeed most subjects did not report any difficulty in identifying the two syllables /pa/ and /ka/. The effect of task must also be taken into account when evaluating multisensory results. Prior work on the McGurk effect sought to investigate the interference caused by visual information on a subject's correct identification of the auditory syllable [Bertelson et al., 2003; Massaro, 1998; Rosenblum and Saldana, 1992; Saldana and Rosenblum, 1993]. Subjects were instructed to concentrate on what they heard to reduce any effect of selective attention to the visual stimuli. In the current study, subjects focused equally on auditory and visual information in order to make a decision about whether the auditory and visual stimuli were congruent. It is possible that this resulted in greater bimodal interaction or a greater contribution from the visual system. These task effects, however, do not account for the new finding of V1 modulation by a change in auditory syllable (Fig. 2, top row). Directing more attention to the visual stimulus would instead be expected to reduce the influence of the auditory syllable on visual processing. In comparing any incongruent stimulus to the baseline of four congruent movies, the effect of differential attention to novel stimuli must also be considered. However, additional comparisons of the incongruent conditions to a congruent change (congruent /pa/ to congruent /ka/) also demonstrated greater BOLD activity for the incongruent conditions. Thus, the effect of attention to a novel stimulus cannot explain the current findings.

The adaptation paradigm may be instrumental in revealing these small signal changes based on percept. In adaptation fMRI, a population of neurons exposed to similar repeated stimuli will gradually decrease their activity, with a subsequent "release from adaptation" when presented with a novel stimulus [Grill-Spector et al., 1999, 2006]. The adaptation effect has been demonstrated to detect differences in retinotopic spatial and orientation specificity [Boynton and Finney, 2003; Grill-Spector et al.,

1999; Murray et al., 2006], object shape and format [Kourtzi and Kanwisher, 2001; Vuilleumier et al., 2005; Winston et al., 2004], identity and face perception [Winston et al., 2004], and higher order discrimination of words and phonetic categories [Jaaskelainen et al., 2004; Raizada and Poldrack, 2007; Wheatley, 2006]. Studies have suggested that the reduced activity seen with repetition or adaptation is due to similarity in physical stimuli [Roberts and Summerfield, 1981; Shigeno, 2002]; however, cross-modal adaptation in language tasks [Buckner et al., 2000; Carlesimo et al., 2003; Jaaskelainen et al., 2004; Kim et al., 2004; McKone and Dennis, 2000; Schacter et al., 2004] implies that adaptation can occur in response to percepts [Carlesimo et al., 2003; Kim et al., 2004] since the visual and auditory stimuli do not share basic sensory properties. If distinct populations of neurons contained within the same imaging voxel respond differentially to percept, this result would be cancelled out in a standard voxel-wise subtraction fMRI analysis [Calvert, 2001; Grill-Spector and Malach, 2001; Grill-Spector et al., 1999, 2006].

These findings have implications for the clinical evaluation of hearing impaired patients. Grant et al. [1998] reported significant benefits for audiovisual stimuli relative to auditory speech alone in hearing impaired subjects. Pre-lingually deaf children with cochlear implants who show the greatest degree of audiovisual benefit also produce more intelligible speech [Lachs et al., 2001]. If hearing loss is chronic or slowly progressive, patients develop an expectation of unreliable auditory information and are more likely to favor the visual modality when audiovisual information is incongruent. Over the long term, this selective neural processing could lead to changes in brain connectivity between multisensory and primary areas. Hearing aids and cochlear implants may be capable of reversing this pattern [Lee et al., 2001, 2003; Roland et al., 2001]. If patients retain cortical plasticity, particularly in areas responsible for the perception of audiovisual speech, they may be expected to achieve greater success with these devices. Pre-operative functional imaging of cochlear implant candidates could be a helpful prognostic tool. Further research in this area is indicated.

ACKNOWLEDGMENTS

The authors thank Don Eddington, Daniel Goldenholtz, Sari Levänen, and Jyrki Ahveninen for helpful comments on the experiments and the manuscript, to Mikko Sams and Riikka Möttönen for stimulus design, and to Dan Wakeman, Deirdre Foxe, Balaji Lakshmanan, and Mary Foley for technical assistance.

REFERENCES

- Bandettini PA, Jesmanowicz A, Van Kylen J, Birn RM, Hyde JS (1998): Functional MRI of brain activation induced by scanner acoustic noise. *Magn Reson Med* 39:410–416.

- Beauchamp MS, Lee KE, Argall B, Martin A (2004): Integration of auditory and visual information about objects in superior temporal sulcus. *Neuron* 41:809–823.
- Benevento L, Fallon J, Davis B, Rezak M (1977): Auditory-visual interaction in single cells in the cortex of the superior temporal sulcus and the orbital frontal cortex of the macaque monkey. *Exp Neurol* 57:849–872.
- Bertelson P, Vroomen J, de Gelder B (2003): Visual recalibration of auditory speech identification: a McGurk after effect. *Psychol Sci* 14:592–597.
- Boynton GM, Finney EM (2003): Orientation-specific adaptation in human visual cortex. *J Neurosci* 23:8781–8787.
- Buckner R, Koutstaal W, Schacter DL, Rosen BR (2000): Functional MRI evidence for a role of frontal and inferior temporal cortex in amodal components of priming. *Brain* 123:620–640.
- Bushara KO, Grafman J, Hallett M (2001): Neural correlates of audio-visual stimulus onset asynchrony detection. *J Neurosci* 21:300–304.
- Callan D, Jones J, Munhall K, Callan A, Kroos C, Vatikiotis-Bateson E (2003): Neural processes underlying perceptual enhancement by visual speech gestures. *Neuroreport* 14:2213–2218.
- Calvert G, Brammer MJ, Bullmore ET, Campbell R, Iversen SD, David AS (1999): Response amplification in sensory-specific cortices during crossmodal binding. *Neuroreport* 10:2619–2623.
- Calvert GA (2001): Crossmodal processing in the human brain: Insights from functional neuroimaging studies. *Cereb Cortex* 11:1110–1123.
- Calvert GA, Campbell R (2003): Reading speech from still and moving faces: The neural substrates of visible speech. *J Cogn Neurosci* 15:57–70.
- Calvert GA, Lewis J (2004): Hemodynamic studies of audiovisual interactions. In: Calvert G, Spence C, Stein B, editors. *The Handbook of Multisensory Processes*. Cambridge, MA: MIT Press. pp 483–502.
- Calvert GA, Bullmore ET, Brammer MJ, Campbell R, Williams SC, McGuire P, Woodruff PW, Iversen SD, David AS (1997): Activation of auditory cortex during silent lipreading. *Science* 276:593–596.
- Calvert GA, Brammer MJ, Bullmore ET, Campbell R, Iversen SD, David AS (1999): Response amplification in sensory-specific cortices during crossmodal binding. *Neuroreport* 10:2619–2623.
- 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.
- Campbell R, MacSweeney M, Surguladze S, Calvert GA, McGuire P, Suckling J, Brammer MJ, David AS (2001): Cortical substrates for the perception of face actions: An fMRI study of the specificity of activation for seen speech and for meaningless lower-face acts (gurning). *Cogn Brain Res* 12:233–243.
- Carlesimo G, Turriziani P, Paulesu E, Gorini A, Caltagirone C, Fazio F, Perani D (2003): Brain activity during intra- and cross-modal priming: New empirical data and review of the literature. *Neuropsychologia* 42:14–24.
- Champoux F, Tremblay C, Mercier C, Lassonde M, Lepore F, Gagne J-P, Theoret H (2006): A role for the inferior colliculus in multisensory speech integration. *Neuroreport* 17:1607–1610.
- Dale A, Fischl B, Sereno M (1999): Cortical surface-based analysis. I. Segmentation and surface reconstruction. *Neuroimage* 9:179–194.
- Dodd B (1977): The role of vision in the perception of speech. *Perception* 6:21–40.
- Edmister W, Talavage T, Ledden P, Weisskoff R (1999): Improved auditory cortex imaging using clustered volume acquisitions. *Hum Brain Mapp* 7:89–97.
- Falchier A, Clavagnier S, Barone P, Kennedy H (2002): Anatomical evidence of multimodal integration in primate striate cortex. *J Neurosci* 22:5749–5759.
- Fingelkurtz AA, Fingelkurtz AA, Krause C, Mottonen R, Sams M (2003): Cortical operational synchrony during audio-visual speech integration. *Brain Lang* 85:297–312.
- Finney EM, Fine I, Dobkins KR (2001): Visual stimuli activate auditory cortex in the deaf. *Nat Neurosci* 4:1171–1173.
- Fischl B, Sereno M, Dale A (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, Ségonne F, Salat DH, Busa E, Seidman LJ, Goldstein J, Kennedy D, Caviness V, Makris N, Rosen B, Dale AM (2004): Automatically parcellating the human cerebral cortex. *Cereb Cortex* 14:11–22.
- Fowler CA (2004): Speech as a supramodal or amodal phenomenon. In: Calvert G, Spence C, Stein B, editors. *The Handbook of Multisensory Processes*. Cambridge, MA: MIT Press. pp 189–201.
- Giard M, Peronnet F (1999): Auditory-visual integration during multimodal object recognition in humans: A behavioral and electrophysiological study. *J Cogn Neurosci* 11:473–490.
- Grant KW, Seitz P (1998): Measures of auditory-visual integration in nonsense syllables and sentences. *J Acoust Soc Am* 104:2438–2450.
- Grant KW, Seitz P (2000): The use of visual speech cues for improving auditory detection of spoken sentences. *J Acoust Soc Am* 108:1197–1208.
- Grant KW, Walden B, Seitz P (1998): Auditory-visual speech recognition by hearing-impaired subjects: Consonant recognition, sentence recognition, and auditory-visual integration. *J Acoust Soc Am* 103(5, Part 1):2677–2690.
- Grill-Spector K, Malach R (2001): fMRI-adaptation: A tool for studying the functional properties of human cortical neurons. *Acta Psychol* 107:293–321.
- Grill-Spector K, Kushnir T, Edelman S, Avidan G, Itzhak Y, Malach R (1999): Differential processing of objects under various viewing conditions in the human lateral occipital cortex. *Neuron* 24:187–203.
- Grill-Spector K, Henson R, Martin A (2006): Repetition and the brain: Neural models of stimulus-specific effects. *Trends Cogn Neurosci* 10:14–23.
- Gusnard DA, Raichle ME (2001): Searching for a baseline: functional imaging and the resting human brain. *Nat Rev Neurosci* 2:685–694.
- Hall DA, Haggard MP, Akeroyd MA, Palmer AR, Summerfield AQ, Elliott MR, Gurney EM, Bowtell RW (1999): “Sparse” temporal sampling in auditory fMRI. *Hum Brain Mapp* 7:213–223.
- Hayes EA, Tiippana K, Nicol TG, Sams M, Kraus N (2003): Integration of heard and seen speech: A factor in learning disabilities in children. *Neurosci Lett* 351:46–50.
- Jaaskelainen IP, Ojanen V, Ahveninen J, Auranen T, Levanen S, Mottonen R, Tarnanen I, Sams M (2004): Adaptation of neuro-magnetic n1 responses to phonetic stimuli by visual speech in humans. *Neuroreport* 15:2741–2744.
- Jones J, Callan D (2003): Brain activity during audiovisual speech perception: An fMRI study of the McGurk effect. *Neuroreport* 14:1129–1133.

- Jones J, Jarick M (2006): Multisensory integration of speech signals: The relationship between space and time. *Exp Brain Res* 174:588–594.
- Kaas J, Hackett T (2000): Subdivisions of auditory cortex and processing streams in primates. *Proc Natl Acad Sci USA* 97:11793–11799.
- Kaas J, Collins C (2004): The resurrection of multisensory cortex in primates: Connection patterns that integrate modalities. In: Calvert G, Spence C, Stein B, editors. *The Handbook of Multisensory Processes*. Cambridge, MA: MIT Press. pp 285–293.
- Kim J, Davis C, Krins P (2004): Amodal processing of visual speech as revealed by priming. *Cognition* 93:B39–B47.
- Kourtzi Z, Kanwisher N (2001): Representation of perceived object shape by the human lateral occipital cortex. *Science* 293:1506–1509.
- Lachs L, Pisoni D, Kirk K (2001): Use of audiovisual information in speech perception by prelingually deaf children with cochlear implants: A first report. *Ear Hear* 22:236–251.
- Lee DS, Lee JS, Oh SH, Kim S-K, Kim J-W, Chung J-K, Lee MC, Kim CS (2001): Cross-modal plasticity and cochlear implants 409:149–150.
- Lee JS, Lee DS, Seung HO, Kim CS, Kim J-W, Hwang CH, Koo J, Kang E, Chung J-K, Myung CL (2003): Pet evidence of neuroplasticity in adult auditory cortex of postlingual deafness. *J Nucl Med* 44:1435–1439.
- Levänen S, Utuela K, Stephan S, Hari R (2001): Cortical representation of sign language: Comparison of deaf signers and hearing non-signers. *Cereb Cortex* 11:506–512.
- MacSweeney M, Calvert GA, Campbell R, McGuire P, David AS, Williams SC, Woll B, Brammer MJ (2002): Speechreading curcuits in people born deaf. *Neuropsychologia* 40:801–807.
- Massaro DW (1998): *Perceiving Talking Faces*. Cambridge, MA: MIT Press. 503 p.
- Massaro DW (2004): From multisensory integration to talking heads and language learning. In: Calvert G, Spence C, Stein B, editors. *The Handbook of Multisensory Processes*. Cambridge, MA: MIT Press. pp 153–176.
- Mazoyer B, Zago L, Mellet E, Bricogne S, Etard O, Houde O, Crivello F, Joliot M, Petit L, Tzourio-Mazoyer N (2001): Cortical networks for working memory and executive functions sustain the conscious resting state in man. *Brain Res Bull* 54:287–298.
- McGurk H, MacDonald J (1976): Hearing lips and seeing voices. *Nature* 264:746–748.
- McKone E, Dennis C (2000): Short-term implicit memory: Visual, auditory, and cross-modality priming. *Psychon Bull Rev* 7:341–346.
- Miller L, D’Esposito M (2005): Perceptual fusion and stimulus coincidence in the cross-modal integration of speech. *J Neurosci* 25:5884–5893.
- Molholm S, Ritter W, Murray MM, Javitt DC, Schroeder CE, Foxe JJ (2002): Multisensory auditory-visual interactions during early sensory processing in humans: A high-density electrical mapping study. *Cogn Brain Res* 14:115–128.
- Möttönen R (2002): Processing of changes in visual speech in the human auditory cortex. *Cogn Brain Res* 13:417–425.
- Munhall K, Vatikiotis-Bateson E (2004): Spatial and temporal constraints on audiovisual speech perception. In: Calvert G, Spence C, Stein B, editors. *The Handbook of Multisensory Processes*. Cambridge, MA: MIT Press. pp 177–188.
- Munhall K, Gribble P, Sacco L, Ward M (1996): Temporal constraints on the McGurk effect. *Percept Psychophys* 58:351–362.
- Murray SO, Olman CA, Kersten D (2006): Spatially specific fMRI repetition effects in human visual cortex. *J Neurophysiol* 95:2439–2445.
- Musacchia G, Sams M, Nicol T, Kraus N (2006): Seeing speech affects acoustic information processing in the human brainstem. *Exp Brain Res* 168:1–10.
- Ojanen V, Mottonen R, Pekkola J, Jaaskelainen IP, Joensuu R, Autti T, Sams M (2005): Processing of audiovisual speech in broca’s area. *Neuroimage* 35:333–338.
- Olson IR, Gatenby JC, Gore JC (2002): A comparison of bound and unbound audio-visual information processing in the human cerebral cortex. *Cog Brain Res* 14:129–138.
- Pekkola J, Ojanen V, Autti T, Jaaskelainen IP, Mottonen R, Tarkiainen A, Sams M (2005): Primary auditory cortex activation by visual speech: An fMRI study at 3t. *Neuroreport* 16:125–128.
- Pekkola J, Laasonen M, Ojanen V, Autti T, Jaaskelainen IP, Kujala T, Sams M (2006): Perception of matching and conflicting audiovisual speech in dyslexic and fluent readers: An fMRI study at 3t. *Neuroimage* 29:797–807.
- Petitto LA, Zatorre RJ, Gauna K, Nikelski EJ, Dostie D, Evans AC (2000): Speech-like cerebral activity in profoundly deaf people processing signed languages: Implications for the neural basis of human language. *Proc Natl Acad Sci USA* 97:13961–13966.
- Raij T, Uutela K, Hari R (2000): Audiovisual integration of letters in the human brain. *Neuron* 28:617–625.
- Raizada RDS, Poldrack RA (2007): Selective amplification of stimulus differences during categorical processing of speech. *Neuron* 56:726–740.
- Roberts M, Summerfield Q (1981): Audiovisual presentation demonstrates that selective adaptation in speech perception is purely auditory. *Percept Psychophys* 30:309–314.
- Roder B, Stock O, Bien S, Neville H, Rosler F (2002): Speech processing activates visual cortex in congenitally blind humans. *Eur J Neurosci* 16:930–936.
- Roland PS, Tobey EA, Devous MD (2001): Preoperative functional assessment of auditory cortex in adult cochlear implant users. *Laryngoscope* 111:77–83.
- Rosenblum LD, Saldana HM (1992): Discrimination tests of visually-influenced syllables. *Percept Psychophys* 52:461–473.
- Saldana HM, Rosenblum LD (1993): Selective adaptation in speech perception using a compelling audiovisual adaptor. *J Acoust Soc Am* 95:3658–3661.
- Sams M, Aulanko R, Hamalainen M, Hari R, Lounasmaa O, Lu S-T, Simola J (1991): Seeing speech: Visual information from lip movements modifies activity in the human auditory cortex. *Neurosci Lett* 127:141–145.
- Sawamura H, Orban GA, Vogels R (2006): Selectivity of neuronal adaptation does not match response selectivity: A single-cell study of the fMRI adaptation paradigm. *Neuron* 49:307–318.
- Schacter DL, Buckner R (1998): On the relations among priming, conscious recollection, and intentional retrieval: Evidence from neuroimaging research. *Neurobiol Learn Mem* 70:284–303.
- Schacter DL, Dobbins IG, Schnyer DM (2004): Specificity of priming: A cognitive neuroscience perspective. *Nature Rev Neurosci* 5:853–862.
- Schroger E, Widmann A (1998): Speeded responses to audiovisual signal changes result from bimodal integration. *Psychophysiology* 35:755–759.
- Shams L, Iwaki S, Chawla A, Bhattacharyya J (2005): Early modulation of visual cortex by sound: An MEG study. *Neurosci Lett* 378:76–81.

- Shigeno S (2002): Anchoring effects in audiovisual speech perception. *J Acoust Soc Am* 111:2853–2861.
- Shulman GL, Corbetta M, Buckner RL, Raichle ME, Fiez JA, Miezin FM, Petersen SE (1997): Top-down modulation of early sensory cortex. *Cereb Cortex* 7:193–206.
- Stein B, Meredith M (1993): *Merging of the Senses*. Cambridge, MA: MIT Press. 224 p.
- Sumby WH, Pollack I (1954): Visual contribution to speech intelligibility in noise. *J Acoust Soc Am* 26:212–215.
- van Atteveldt N, Formisano E, Blomert L, Goebel R (2004): Integration of letters and speech sounds in the human brain. *Neuron* 43:271–282.
- Vuilleumier P, Schwartz S, Duhoux S, Dolan R, Driver J (2005): Selective attention modulates neural substrates of repetition priming and “Implicit” visual memory: Suppressions and enhancements revealed by fMRI. *J Cogn Neurosci* 17:1245–1260.
- Wallace M, Meredith M, Stein B (1993): Converging influences from visual, auditory, and somatosensory cortices onto output neurons of the superior colliculus. *J Neurophysiol* 69:1797–1809.
- Watkins S, Shams L, Tanaka S, Haynes J, Rees G (2006): Sound alters activity in human v1 in association with illusory visual perception. *Neuroimage* 31:1247–1256.
- Wheatley T, Weisberg J, Beauchamp M, Martin A (2005): Automatic priming of semantically related words reduces activity in the fusiform gyrus. *J Cogn Neurosci* 17:1871–1885.
- Wiggs CL, Martin A (1998): Properties and mechanisms of perceptual priming. *Curr Opin Neurobiol* 8:227–233.
- Winston J, Henson R, Fine-Goulden M, Dolan R (2004): fMRI-adaptation reveals dissociable neural representations of identity and expression in face perception. *J Neurophysiol* 92:1830–1839.
- Xu H, Dayan P, Lipkin RM, Qian N (2008): Adaptation across the cortical hierarchy: low-level curve adaptation affects high-level facial-expression judgements. *J Neurosci* 28:3374–3383.