

Research report

A comparison of bound and unbound audio–visual information processing in the human cerebral cortex

Ingrid R. Olson*, J. Christopher Gatenby, John C. Gore

Department of Diagnostic Radiology, Yale School of Medicine, Fitkin Bldg, Rm F14, 333 Cedar Street, New Haven, CT 06510, USA

Abstract

Human speech has auditory (heard speech) and visual (seen speech) qualities. The neural representation of audiovisual integration in speech was investigated using functional magnetic resonance imaging (fMRI). Ten subjects were imaged while viewing a face in four different conditions: with speech and mouth movements synchronized, with speech and mouth movements desynchronized, during silent speech, or while viewing a static face. Subtractions of the different sets of images showed that lipreading primarily activated the STG/STS. Synchronized audio–visual speech and desynchronized audio–visual speech activated similar areas. Regions activated more in the synchronized versus the desynchronized conditions were considered to be those involved in cross-modal integration. One dominant activation focus was found near the left claustrum, a subcortical region. A region-of-interest analysis of the STS and parietal areas found no difference between audio–visual conditions. However, this analysis found that synchronized audio–visual stimuli led to a higher signal change in the claustrum region. This study extends previous results, using other sensory combinations, and other tasks, indicating involvement of the claustrum in sensory integration. © 2002 Elsevier Science B.V. All rights reserved.

Theme: Other systems of the CNS

Topic: Association cortex and thalamocortical relations

Keywords: Claustrum; Cross-modal integration; Sensory integration; Lipreading; Putamen; McGurk effect

1. Introduction

How do the senses converge to create multimodal representations of stimuli? Two dominant theories exist for how sensory input is integrated. Theory 1 is a *site-specific* integration model; it suggests that special purpose regions of cortex process *specific combinations* of sensory input. This theory is based on early neuroanatomical and neurophysiological research which suggested that there were ‘association’ areas of the brain that only processed specific stimulus combinations [44]. These ‘poly-modal’ areas were suggested to be in the arcuate sulcus, superior temporal sulcus (STS), inferior and posterior parietal lobules, the amygdaloid complex including the rhinal cortex and hippocampus, and the superior colliculus (reviewed by Ref. [20]). Theory 2 is a *communication relay* model. This model suggests that neural areas that process *single* sensory stimuli (e.g. unimodal stimuli) are the same

areas that process *multi-sensory* stimuli. Importantly, each sense can access other senses through a subcortical relay area. This theory is primarily supported by negative evidence in that numerous studies have shown that lesions to putative poly-modal areas fail to disrupt sensory integration [20,21].

Human research has studied audio–visual (AV) and tactile–visual integration with functional magnetic resonance imaging (fMRI) and positron emission tomography (PET). Three neural areas have emerged as candidate cross-modal regions: STS, the insula and nearby subcortical structures, and the parietal lobe. The STS has been suggested as a site of AV integration. Calvert et al. [11] indirectly studied AV integration using a lipreading task. They found that silent lipreading activated a number of regions, including the superior temporal gyrus (STG)/STS, a region that partially overlaps with spoken language perception areas. The authors suggested that this region was important for AV integration since it was highly responsive to both visual and auditory language. Thus their findings tend to support the site-specific theory of sensory integration. A more direct study of AV integration com-

*Corresponding author. Tel.: +1-203-785-5052; fax: +1-203-785-6534.

E-mail address: ingrid.olson@yale.edu (I.R. Olson).

pared the neural response to AV stimulation to a unimodal auditory or a unimodal visual stimulation [14]. The results failed to support the STS/STG cross-modal hypothesis since increased activity in response to cross-modal stimulation was found only in visual motion cortex (MT/V5) and primary and secondary auditory cortex. These results failed to confirm the results of their earlier study. The results also tend to support theory 2 since AV integration was found to occur in areas that process unimodal visual and unimodal auditory stimuli. A follow-up study by the same group examined AV integration in speech and found a different pattern of results: activations in the STS and other regions. Non-speech AV integration activated the STS, intraparietal sulcus (IPS), insula/claustum, and the superior colliculi [9,10,13]. These results support theory 1 since the reported activations were in areas that have been previously suggested as cross-modal integration areas and also since the STS and the IPS are areas that are generally thought of as higher order, association areas.

Tactile–visual integration studies have reported a different pattern of results. Hadjikhani and Roland [23] used PET to measure tactile–visual matching. The right claustrum was the only brain area that was activated in the crossmodal condition. Another PET study using a tactile–visual matching paradigm also found effects in the left claustrum in addition to other regions such as the inferior parietal lobes [4]. A similar finding was reported by Horster et al. [25]. They tested tactile–visual integration in monkeys and reported 2-deoxyglucose (2-DG) labeling in the left and right claustrum. The claustrum is a small subcortical area of gray matter, medial to the insula and lateral to the basal ganglia, which contains orderly maps of visual, auditory, and somatosensory cortices. Ettliger and Wilson [20] reviewed the sensory integration literature and concluded that the claustrum was an area that might relay information between the senses. Thus the findings from tactile–visual integration tend to support theory 2, that is, that the senses access each other through a common communication relay area. The generality of this finding is difficult to assess since the three studies that report claustrum activations used similar stimuli (e.g. tactile–visual), and an identical paradigm (e.g. matching). Thus it is difficult to make statements about the involvement of the claustrum in all types of sensory integration.

Other studies have not directly addressed the two theories of sensory integration, but are important to note nevertheless. For instance, Lewis et al. [30] had subjects perform separate visual and auditory motion discrimination tasks while measuring changes in the hemodynamic response (e.g. fMRI). Areas activated by both tasks included lateral parietal cortex, extending into the intraparietal sulcus (IPS), and the insula. A second experiment required an explicit cross-modal speed comparison and found that the cross-modal task activated sites including the IPS and the insula. The parietal activations may have overlapped with regions containing cells that respond to both auditory and visual stimuli, previously identified in animals [2].

In this study we examine the two competing hypotheses about AV integration. Does AV integration occur by using a dedicated polysensory module, such as portions of the parietal lobe or the STS (e.g. theory 1)? Or does AV integration utilize the same neural mechanism as other forms of cross-modal integration, such as tactile–visual, which appear to rely on the claustrum and nearby structures (e.g. theory 2)? We investigated the cortical activation patterns of subjects integrating AV stimuli using fMRI. We compared two different cross-modal conditions: one in which the AV signals were temporally synchronized, and one in which they were desynchronized. Past research has found that temporal or spatial synchrony of two stimulus sources is critical for achieving sensory integration [32,34,42]. By manipulating only signal synchrony, we could use both auditory and visual stimuli in the two comparative conditions. Two additional visual-only conditions, a static face and a silently talking face, allowed us to compare AV integration to silent lipreading. To test the region-specific hypotheses, regions of interest were defined in candidate cross-modal regions, as identified in previous studies: the claustrum, the parietal lobes, and the STS.

2. Materials and methods

2.1. Subjects

Ten volunteers (aged between 18 and 41 years, mean age 27 years) participated in the study. All of the subjects gave informed consent to a protocol reviewed and approved by the Human Investigation Subject Committee of the School of Medicine at Yale University. None had previous or present history of medical illness, and all subjects were right-handed.

2.2. Stimuli

Sequences of MR images were acquired during four different conditions: static visual face (static V), moving visual face (moving V), synchronized audio–visual (synchronized AV), and desynchronized audio–visual (desynchronized AV).

The stimulus in all tasks was a female face, in color, on a gray background. In the static V task, a single video frame of the face was continuously present during the duration of the block. In the moving V task, the face mouthed words at a rate of one every 3 s. In the synchronized AV task, the moving face was synchronized with a soundtrack so that the face appeared to be saying a word every 3 s. We used the ‘McGurk effect’ to manipulate AV integration [33]. In this phenomenon, an audible syllable or word (e.g. ‘bat’) is dubbed onto a videotape of a person mouthing a different syllable or word (e.g. ‘vet’). Subjects typically report hearing something else (e.g. ‘vat’). The McGurk effect was used because it provides compelling evidence of AV integration in speech percep-

tion. The audio track contained the words: bat, bent, boat, might, mail, mat, and moo. The video track contained the words: vet, vest, vow, die, deal, dead, and goo. The McGurk effect gave rise to the percept of the following words when the stimuli were temporally synchronized, in order: vat, vent, vote, night, nail, gnat, and new [5]. This McGurk condition was contrasted to a condition in which the AV signals were temporally desynchronized by 1 s so that the audio stimulus was heard correctly, and thus, no McGurk effect is produced. The same moving face had an audio track superimposed between mouth movements, as in a poorly dubbed movie. A different set of audio stimuli was used to prevent the audio component of the synchronized AV condition from becoming unduly salient, due to repetition, and perhaps creating a bias to perceive unfused words. Word lists were equated for frequency of usage. The audio track contained the words: vein, best, mate, doe, meal, bead, and boo; the video track was identical to the synchronized AV video track (Fig. 1).

The actress was videotaped while speaking these words; she was used for all stimuli. The stimuli were digitized using i-Movie video software, and were saved in Quicktime format. Sounds were recorded in a soundproof booth and dubbed onto the videos using SoundEdit software. All movies were presented from a Macintosh computer using Psyscope software [16].

In the static V condition, the same face remained on the screen for 21 s. In the moving V, synchronized AV, and desynchronized AV conditions, each stimulus was 3 s long. Seven of these stimuli were grouped into 21-s blocks, followed by a 6-s break, containing a 1-s fixation cross. Each condition was presented three times per run, in pseudo-random order, for a total run time of 5.5 min. Each subject was scanned four times (some subjects were

scanned up to six times, but their last runs were discarded to equate the total number of runs across subjects).

2.3. Task

Subjects were instructed to pay attention to the mouth region of the actor and to try to understand what she was saying, in the synchronized AV, desynchronized AV, and moving V conditions. They were instructed to say the word ‘one’ to themselves during the static V condition to control for subvocalization and also to focus attention [11]. Subjects were also instructed to press a key at the end of each block, during the 6-s rest interval, to ensure that they were alert.

Visual stimuli were projected onto a screen located at the base of the scanner bed via an LCD projector (resolution 640×480 pixels). The stimuli were viewed through a mirror angled above the subject’s head in the scanner. Auditory stimuli were presented through an audio headset. The subject set the sound to a comfortable level prior to acquiring fMRI data.

After scanning, subjects performed two short behavioral tests. The first test was a forced-choice assessment of their memory for the visual and auditory words presented during scanning. This test included a test of memory for AV fusions (e.g. McGurk percepts) allowing assessment of the robustness of AV fusion. Subjects were presented with two words on a computer monitor, one of which had occurred during the experiment. Word frequency, usage, and phonemic structure were similar across word lists. They were instructed to choose the word that was most familiar by hitting a designated computer key.

The second test assessed word perception. Subjects viewed the desynchronized AV and the synchronized AV

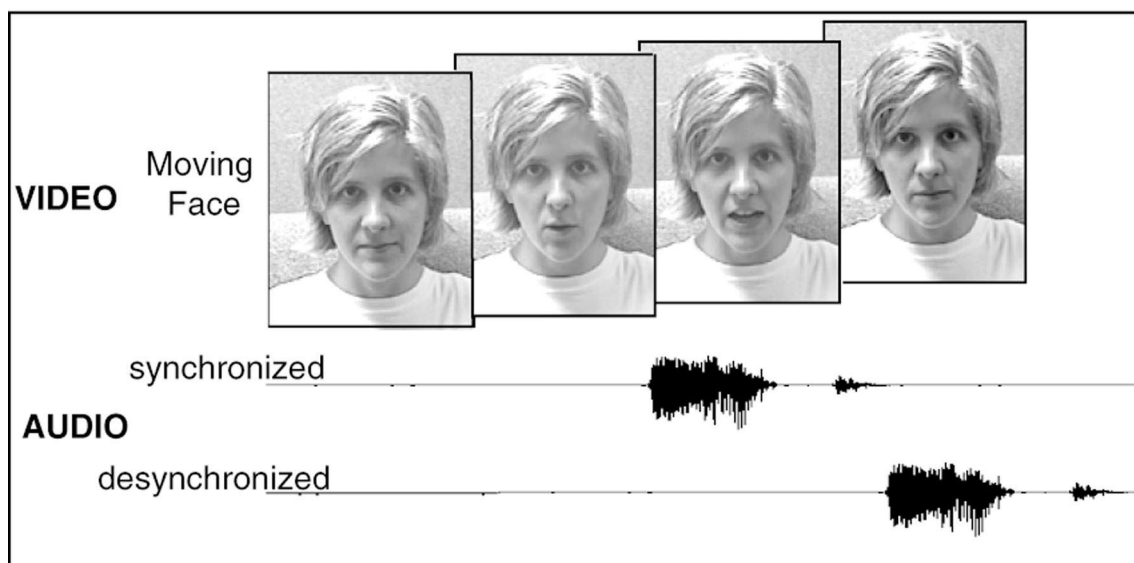


Fig. 1. A schematic illustration of the synchronized AV and desynchronized AV conditions. In the synchronized AV condition, the onset of the speech signal was synchronized with the onset of the mouth movements. In the desynchronized AV condition, the onset of the speech signal was desynchronized from the onset of the mouth movement.

stimuli. They were instructed to verbally report what they heard the actress say. The stimuli were shown under two different audio conditions, background noise (to simulate the scanning environment) and no background noise. Performance was collapsed across conditions.

2.4. Imaging and image analysis

2.4.1. Scanning procedure

MRI scans were acquired using a 1.5 T General Electric Signa Lx scanner. Subjects' heads were immobilized within a circularly polarized head coil. Anatomical images in the sagittal and coronal planes (TR=500 ms; TE=14 ms) were acquired prior to functional images, to aid in normalization of individual subject data into standard stereotactic space [43]. Functional images were acquired using echo planar imaging (gradient echo single shot sequence, 220 images per slice, FOV=20×20 cm, matrix=64×64, NEX=1, TR=1500 ms, TE=60 ms, flip angle=60°). Fourteen 8-mm-thick coronal slices, (gap interval=1 mm), aligned parallel to the AC–PC line, covered most of the visual and temporal cortices, and part of the frontal cortex. Voxels were 3.1×3.1×8 mm³.

2.4.2. Data analysis—artifact removal

Before data analysis, all functional images were screened for obvious artifacts by looking at variations in parameters including the center of mass. Images that showed visible motion or other artifacts (such as ghosting) were removed. Next, images were motion corrected for three translation directions and three possible rotations using the SPM-99 software (Wellcome Department of Cognitive Neurology, London, UK). Maps were spatially smoothed using a Gaussian filter with a full-width half-maximum value of two pixels in plane.

2.4.3. Data analysis—single subject activation maps

Data analysis was performed using software written in MATLAB (MathWorks, MA). Maps of *t*-values were created, corrected for linear drift in the signal by comparing images between two conditions. The *t*-tests were calculated for each series of images separately. This analysis used a hemodynamic lag of 3 s, thereby removing effects of previous conditions. The individual maps were transformed into standardized Talairach atlas coordinates [43] using eight anatomical anchor points (AC, PC, and the superior, inferior, anterior, posterior, left and right most points on the cortical surface).

2.4.4. Data analysis—composite maps

The standardized statistical maps were used in standard linear contrasts, which represented mean activation levels between one condition and another condition. *t*-statistics were computed and used to derive a measure of the signal change at each voxel relative to its own intrinsic noise variability. Next a standard linear contrast was computed

for each complex effect. This procedure generates a single value for each voxel that was determined by the weighted comparison of one (or more) tasks against another set of tasks. Under the null hypothesis of no effect, the expected value of this contrast is equal to zero. The extent to which the contrast value reliably deviates from zero is then assessed. Significance levels ($\alpha=0.05$) were assessed using a nonparametric randomization test. The randomization procedure creates a population distribution for each voxel by repeatedly calculating the value of the contrast when the *t*-values of half the subjects, randomly chosen, have a reversed sign (e.g. + or -). This randomization was performed 500 times, generating a sampling distribution of the linear contrast measures. The linear contrast measure, calculated without sign reversal, was assigned a *P*-value based on its position in this distribution. The contrast maps were overlaid on a composite anatomical image.

2.4.5. Data analysis—region of interest analysis

Regions of interest analysis (ROI) were defined in standard Talairach space and overlaid on individual maps transformed to the Talairach space. Three ROIs were defined: the claustrum/putamen, the STS, and the medial to posterior portions of the parietal lobe. All ROIs were defined anatomically by the atlas of Talairach and Tournoux [43], with the caveat that they included the centroid of cross-modal activation as identified in previous neuro-imaging studies.

The claustrum ROI was medial to the insula, lateral to the thalamus, and superior to the hippocampal/amygdaloid regions. The claustrum/insula activation in Hadjikhani and Roland [23] was at -32, -9, 13 in the right hemisphere. The claustrum/insula activation in Banati et al. [4] was more anterior and inferior (-36, 4, 4) and in the left hemisphere. The claustrum ROI used in our study included the areas found by Hadjikhani and Roland and Banati et al. The center of the ROI in the *x*-direction was ± 29 . In the *y*-direction, the ROI began at 4 and extended posteriorly to -28. The center of the ROI in the *z*-direction was +4.

The STS activation noted by Calvert et al. [13] was centered at -49, -50, 9 in the left hemisphere. A more anterior temporal activation, in the middle and superior temporal gyri was found by Banati et al. [4] at 34, -36, 12 on the right, and -42, -44, 8 on the left. The STS ROI used in our study included the areas found by Calvert et al. [13] and Banati et al. [4] in both hemispheres. Both upper and lower banks were included. The centroids of the ROI were: ± 52 , -43, 7.

The parietal activation in Lewis et al. [30] was centered at 35, -46, 47 in the right hemisphere, and -41, -40, and 47 in the left hemisphere. The parietal activation to tactile-visual integration in Banati et al. [4] was more inferior and anterior (centered at 50, -26, 32, in the right hemisphere and -50, -28, 28, in the left hemisphere). Because the activation in Banati et al. [4] appeared to be in the

somatosensory cortex, it was excluded from our ROI. The parietal ROI was large and consisted primarily of Brodman Area 7 and some of Area 40. The ROI was centered around ± 35 , -55 , 47 .

All ROIs represent reasonable volumes in terms of the known anatomical specifications of these regions. In all ROIs, percent signal change was compared in the synchronized and desynchronized AV conditions, in each hemisphere, using a two-factor repeated measures ANOVA across subjects.

2.4.6. Data analysis—*anatomical localization*

Anatomical designations were based on the atlases of Talairach and Tournoux [43] and Duvernoy [18].

3. Results

3.1. Behavioral

In a forced choice memory task given after the experiment, subjects recognized 83% of the AV fused ‘McGurk’ stimuli (e.g. vat, not vet), suggesting that most subjects were susceptible to the McGurk effect. This was further tested in a perceptual post-test of the McGurk effect. Average correct report of AV fusions was 65%. The AV fusion rate is similar to that obtained by previous investigators [5].

Memory performance was 75% for the words used in the other conditions. In a perceptual post-test, subjects were asked to identify, by speaking aloud, the audio words in the desynchronized AV condition. An average of 93% of the auditory words were correctly identified. Next, subjects were also asked to identify the visual words in the desynchronized AV condition. Subjects found it difficult to accurately report the visual words (19%). These results suggest that (1) subjects fused the audio and visual stimuli in the synchronized AV condition; (2) subjects did not fuse the audio and visual in the desynchronized AV condition; (3) the auditory stimulus was dominant in the desynchronized AV condition; and (4) subjects were attentive to the correct aspects of the stimuli.

3.2. Regional fMRI signal changes

3.2.1. Hemodynamic changes during lipreading

To investigate lipreading, we compared the moving V to the static V condition. Activated voxels ($P < 0.05$; see Fig. 2) were found in the right homologue of Broca’s area, (hereafter referred to as right Broca’s area) and left STG and right STS/STG. At a higher threshold ($P < 0.01$) only the left and right STG remained significant. Previous neuroimaging studies have reported both unilateral and bilateral Broca’s area activations associated with a number of different language tasks including syntactic processing [19], phonological processing of words or letters (for a

review see Ref. [40]), and verbal short-term memory [22]. Broca’s area activations have also been reported in imitation tasks, possibly reflecting a link between motor-production and language perception and production [27,35]. Activated voxels were also found bilaterally in the sensory-motor regions surrounding the central sulcus. Activations in motor areas are often found in language tasks, possibly reflecting subvocalization, or the strong ties of language to motor processing [31]. Additionally, small areas of activation were found in the left basal ganglia, and left anterior cingulate. The anterior cingulate has been postulated to be important in attention, motor modulation, and response selection [15,38]. Our task did not involve a response thus the activations most likely reflect attentional processing. Lipreading was a demanding task, and one would expect a high need for attention. The Talairach coordinates [43] and activation measures are listed in Table 1.

3.2.2. Hemodynamic changes during cross-modal integration

To investigate cross-modal integration, activation patterns in the synchronized AV and desynchronized AV conditions were compared to those obtained in the static V condition. The activation patterns were very similar in these comparisons, with task-dependent signal changes appearing in left inferior frontal gyrus, left Broca’s area, and a large bilateral STG activation extending posteriorly to the STS. Most of these areas are known to be important in various aspects of speech perception and production [8]. The common activation in both comparisons may be due to the presence of similar auditory stimuli in both conditions. Bilateral activations in STG have been reported in many other studies of speech perception [24]. This activation was large and extended medially into parts of MTG, the insula, the claustrum, and basal ganglia. Smaller activations were found bilaterally in the precentral gyrus, and in the left thalamus (Fig. 2). At a higher threshold ($P < 0.001$) only the STG and STS activations were present. Since the patterns were so similar, the Talairach coordinates and activation sizes are listed only for the synchronized AV condition in Table 1.

The contrast between the synchronized AV and desynchronized AV conditions showed one primary activation ($P < 0.05$) in the left claustrum region (Fig. 3). Inspection of individual subjects indicates that the activation was somewhat variable, extending laterally into the insula in some subjects, and more medially into the putamen in others. An additional small activation was found in the left temporal pole (Fig. 3, $y = 4$).

Individual subject activations to the synchronized AV and desynchronized AV conditions were examined in the three ROIs: the claustrum region, portions of the parietal lobe, and the STS. The percent signal change in each ROI was analyzed using a two-factor repeated measures ANOVA across subjects with condition (2) and hemisphere

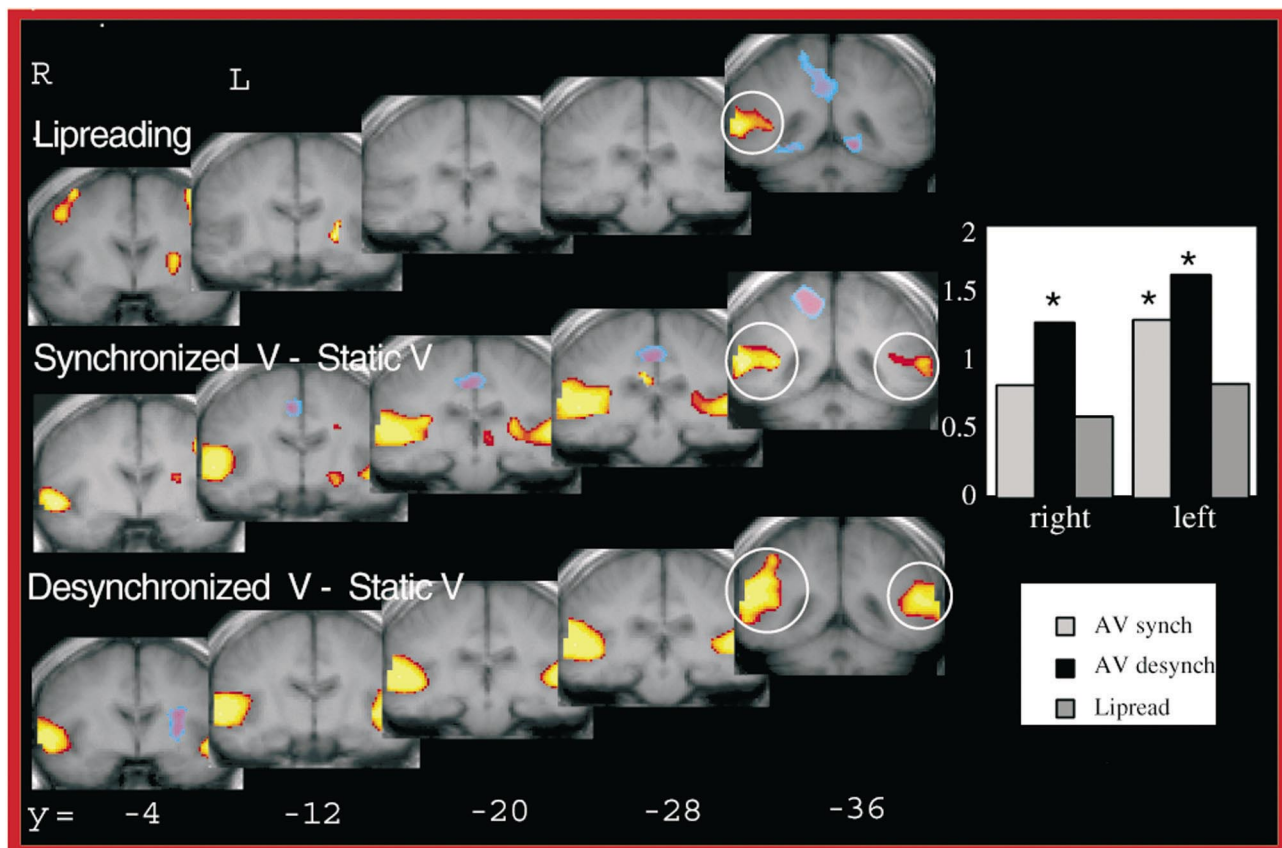


Fig. 2. Activation maps from group contrasts for lipreading (moving V–static V), and the synchronized (synchronized AV–static V) and desynchronized (desynchronized AV–static V) conditions, superimposed on a composite anatomical scan. In this and all other figures, yellow and red colors designate activations. Deactivations are in blue and purple. Yellow, activations of $\sim P=0.01$; red, activations between $P=0.05$ and 0.01 ; blue, $\sim P=0.05$ – 0.01 ; purple, $P=0.01$. The circles enclose the STS/STG region. The graph shows data from an ROI analysis of the STS region; percent signal change is on the y-axis. From left to right, columns indicate: synchronized AV–still; desynchronized AV–still; and moving V–still. The stars indicate comparisons in which there was a significant difference from the moving V–still comparison.

(2) as factors. For the claustrum ROI, there was no effect of condition, ($F(1,9)=2.88$, $P<0.124$), or hemisphere ($F<1$), but there was a significant interaction between condition and hemisphere ($F(1,9)=10.98$, $P<0.009$), suggesting that the conditions differentially affected signal change in the two hemispheres.

The interaction was further investigated by post-hoc t -tests. The first t -test showed that the synchronized condition had a higher signal change than the desynchronized condition in both hemispheres (right: $t(9)=2.60$, $P<0.029$; left: $t(9)=7.28$, $P<0.0001$), although only the left hemisphere remained significant after Bonferroni corrections. The second t -test compared signal change in the left hemisphere to signal change in the right hemisphere, in the synchronized condition. There was significantly higher signal change in the left hemisphere as compared to the right ($t(9)=3.75$, $P<0.005$) after Bonferroni corrections. The percent signal change was higher in the synchronized AV condition compared to the desynchronized AV condition in the left hemisphere of eight of the 10 subjects.

There were no significant effects in the parietal ROI (condition: $F(1,9)=3.05$, $P=0.115$; hemisphere: $F(1,9)=$

1.41 , $P=0.265$; interaction: $F<1$; n.s.) or in the STS ROI (all F s <1 , n.s.). However, in a post-hoc analysis of the STS, we found that the synchronized and desynchronized conditions showed a higher signal change than the lipreading condition, in the left STS ($t(9)=3.88$, $P<0.001$; $t(9)=6.68$, $P<0.0001$). This effect was also found in the right STS for the desynchronized condition ($t(9)=5.68$, $P<0.0001$). This suggests the possibility that the neurons in the STS do crossmodal processing but they do not require temporal synchrony of the AV stimuli for excitation/inhibition.

4. Discussion

4.1. Audiovisual integration

One theory of sensory integration has argued that dedicated cross-modal regions do not exist in the brain, but rather that the senses directly access each other from their sensory-specific unimodal systems via subcortical relay areas [20]. Our study finds evidence in support of this

Table 1

Activation centroids in Talairach coordinates (x, y, z). Size is given in number of activated voxels

Region	x	y	z	Size
Synchronized AV–static V				
R inferior frontal G	48	22	25	69
L Broca's area	-52	13	10	193
L STG	-52	13	-10	443
R STG	48	13	-11	473
R precentral G	49	4	39	97
L precentral G	-56	4	37	124
L sup. basal g.	-25	-4	13	18
L inf. basal g.	-21	-12	-8	206
L medial thalamus	-7	-19	3	31
R STS/STG	58	-38	9	76
L STS/STG	-59	-38	6	102
Moving V–static V				
Brodman 44/45	41	13	27	49
L precentral G	-48	4	52	37
R precentral G	45	4	32	42
L cingulate	-6	4	52	46
L STG	-53	-4	-5	20
L claustrum/basal g.	-26	-4	7	9
R STS/STG	58	-38	10	57
Synchronized AV–desynchronized AV				
L temporal pole	-38	4	-19	19
L claustrum/putamen	-28	-19	1	133

R, right; L, left; sup, superior; inf, inferior; MTG, middle temporal gyrus; STG, superior temporal gyrus; STS, superior temporal sulcus; G, gyrus; g, ganglia.

theory (e.g. theory 2) because synchronized audio–visual stimuli activated many of the same sensory-specific areas of the brain as did desynchronized audio–visual stimuli, suggesting that these areas are important in both unimodal and multimodal processing. Both conditions activated classic speech perception areas. However, integrated sight and sound additionally activated the claustrum/putamen, an area medial to the insula. This was shown in the activation pattern and in the ROI analysis (Fig. 3). It has been previously noted that the claustrum may be a potential ‘relay’ area for the senses [20,23].

This result extends previous PET findings showing that the claustrum is activated in tactile–visual integration [4,23]. Also, a 2-DG study in monkeys found that the claustrum was the most highly labeled region in a tactile–visual cross-modal task [25]. Our study's claustrum activation was more ventromedial compared to that found by Hakjikhani et al. [23], and was in the left, not the right, hemisphere. Banati et al. [4] reported claustrum activations that were more superior to ours but in the same hemisphere. Horster et al. [25] tested tactile–visual integration in monkeys and reported 2-DG labeling in the left and right claustrum, but greater in the left. The claustrum contains orderly maps of visual, auditory, and somato-sensory cortices [36]. It is possible that the difference in activation patterns found between our task and others is due to the activation of different topographically organized maps. Linguistic stimuli may receive greater processing by

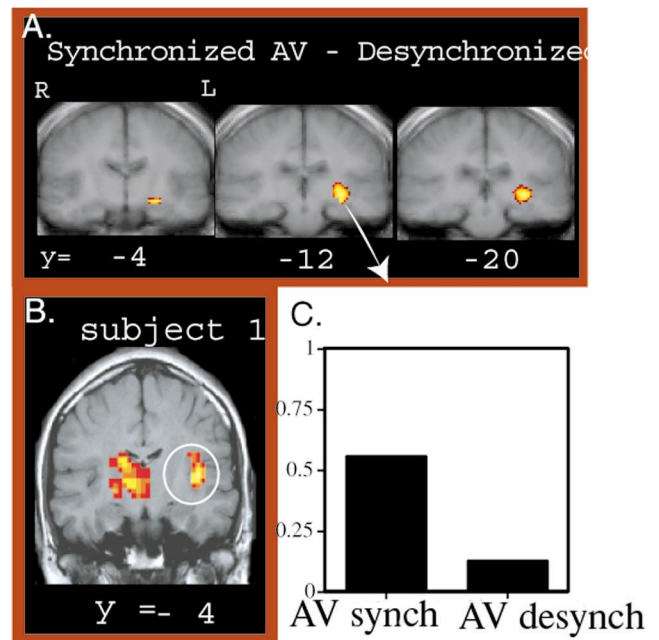


Fig. 3. (a) Activation maps from group contrasts, superimposed on a composite anatomical scan. Cross-modal integration activated a small area in the temporal pole and the claustrum/putamen. Yellow and red colors designate activations. Deactivations are in blue and purple. Yellow, activations of $\sim P=0.01$; red, activations between $P=0.05$ and 0.01 ; blue, $\sim P=0.05-0.01$; purple, $P=0.01$. (b) Activation maps from a single subject, superimposed on their anatomical scans, to show cross-modal integration activations in the right claustrum/putamen. The circle encloses this region. (c) ROI analysis showing percent signal change on the y -axis. There was a significant difference between conditions on both the left and right sides (left side shown).

the left hemisphere and use the left claustrum/putamen to a greater degree than the right.

The ROI analysis also showed that both the synchronized and desynchronized AV conditions activated the STS to a greater degree than the unimodal lipreading condition. Because both conditions activated the STS more strongly, this suggests that the STS was not participating in the type of cross-modal binding that occurs in a time-locked, temporally synchronized manner. One possibility is that neurons in the STS process both auditory and visual signals but do not require temporal synchrony of the AV stimuli for excitation/inhibition. Bushara et al. [7] used PET to image subjects while they performed an AV task. Their findings suggest that subcortical areas such as the insular region are important in detecting temporal synchrony of cross-modal stimuli. Thus subcortical areas, but not cortical areas such as the STS, may be important in detecting temporal synchrony of signals arising from divergent sensory sources.

Another possibility is that cells in the human STS are organized in a mosaic pattern. Thus cells responsive to unimodal auditory, unimodal visual, and bimodal stimuli are intermixed. If true, the activation patterns to these different stimulus types would look similar but possibly

exhibit quantitative differences. And indeed, all comparisons showed activations in the STS but to a different degree. The synchronized AV condition may have activated cells responsive to bimodal stimuli while the desynchronized condition activated cells responsive to two types of unimodal stimuli. The lipreading condition may have activated cells only responsive to visual stimuli.

The audio–visual integration activation extended into the putamen. The putamen is classically thought to constitute the motor input stage of the basal ganglia. Functional knowledge of the human putamen derives primarily from Parkinson's disease, which is caused by degeneration of dopaminergic neurons in the putamen and globus pallidus. The dominant symptoms in Parkinson's are impaired initiation of movement, slowness of movement, and tremor. However, patients with Parkinson's disease also have cognitive deficits in timing. Patients have difficulty recognizing the temporal separation of two stimuli presented in the auditory, visual, and tactile domains [3,26]. Thus degeneration of the putamen and related structures leads to a larger allowable temporal mismatch in the perception of temporal synchrony, suggesting that these structures register stimulus co-occurrence. A comparison between patients with damage to the neocortex and patients with Parkinson's disease showed that only the latter patients exhibited impaired perception of temporal synchrony [29]. Thus if the claustrum acts as a relay station for signals from different modalities, the putamen may act as a timer, signaling that two stimuli occurred simultaneously. Bushara et al. [7] suggest that a different subcortical structure is important in cross-modal temporal synchrony detection: the insula. They used PET to examine the detection of temporal synchrony in AV stimuli. They found that a network of areas was activated by this task, but the area most consistently activated was the right insula.

Many neuroimaging studies of sensory integration have compared cross-modal to unimodal conditions (but see Refs. [7,13]). Our study compared conditions that contained both visual and auditory signals, but with different temporal patterns. Temporal synchrony of stimuli is a major determinant of integration [12,34,41]. In a review of the literature, Summerfield suggested that temporal co-registration of various stimulus features was the binding factor in sensory integration [42].

4.2. Silent lipreading

The STS was activated in the silent lipreading condition, as compared to perception of a static face, replicating previous results [11]. The region of activation in our task was similar to that activated by heard speech in the synchronized AV and desynchronized AV conditions, especially in the right hemisphere. However, as noted earlier, the ROI analysis showed that the activation to lipreading was not as strong as it was to audio–visual inputs, either synchronized or desynchronized.

4.3. Limitations of the present study

We believe that the claustrum/putamen activations found in our study reflects audio–visual integration, however it is possible that these activations reflect a more specific process of audio–visual *speech* integration. Future research will need to address this question by using non-speech stimuli. In light of previous research suggesting a role for the claustrum in tactile–visual integration, it seems unlikely that linguistic stimuli would uniquely activate these regions.

It is also possible that registration errors created by averaging across individual subjects' brains modified signal localization. Other researchers have reported that cross-modal tasks activate the insula, a region that is directly lateral to the claustrum. It is possible that what we are terming the 'claustrum region' is actually part of the insula, which was misregistered during averaging.

4.4. Lack of differential activations in the parietal lobe

Parietal lobe activations were not found in any of the analyses conducted in this study. This is not surprising when placed in the context of previous studies of cross-modal processing that have only inconsistently reported activations in parietal lobes. Even when activations are found in this region, different interpretations of the data have been employed. Lewis et al. [30] suggest that the IPS activations to AV motion reported in their paper could be due to response selection [2,28] or attentional modulation rather than to crossmodal binding per se. It is possible that our task did not activate this region because we did not have a motor response and the attentional requirements were similar across conditions.

4.5. Lack of differential activations to cross-modal stimuli in the STS

The upper bank of the monkey STS contains a polysensory area (reviewed in Ref. [17]), hence we expected to see activation of a similar region of the human STS during perception of the McGurk illusion. One recent fMRI study of audio–visual integration compared the neural response to a talking face to unimodal auditory (speech) or unimodal visual (face) stimulation [14] and found that the AV stimulation resulted in an increased hemodynamic response in unimodal sensory cortices, but not in STS/STG. However, another study, using similar stimuli, found that the STS/STG was activated more to a talking face [13]. In our task the STS/STG region was strongly activated during cross-modal integration in the synchronized AV condition, but it was equally activated during the desynchronized AV condition, suggesting that this region participates equally in the processing of auditory and visual speech stimuli but not in their supra-modal integration. The STS was activated more strongly in both AV conditions as compared to

the unimodal lipreading condition. It is possible that neurons in human STS/STG process cross-modal information but this information is not coded in the peak firing rate and subsequent oxygen consumption as measured by fMRI. This information could be coded as increased coherence between neurons, which would not be more readily assessed electrophysiologically. However, a large body of research from monkeys and cats suggests that cross-modal information is encoded by changes in peak firing rate.

Neuroimaging studies have implicated the human STS in the perception of hand, eye, mouth, and whole body movements (reviewed in Ref. [1]). Similarly, some neurons in the STS of the monkey respond selectively to the perception of body movements [6,37,39]. Salient biological motion, such as linguistic mouth movements or eye-movements, with communicative value, may particularly activate this region. Taken together, existing data suggest that the STS/STG region may be important for the analysis of biological motion of audio or visual stimuli rather than for cross-modal integration.

Acknowledgements

This work was supported by NIH NS33332 to J.C. Gore. We thank Cheryl Lacadie, Terry Hickey, and Pawel Skudlarski for technical assistance. We also thank Truett Allison for helpful discussions, and Lisa Suatoni, Steven Frost, and Ainer Mencil for help in stimulus preparation.

References

- [1] T. Allison, A. Puce, G. McCarthy, Social perception from visual cues: role of the STS region, *Trends Cogn. Sci.* 4 (2000) 258–267.
- [2] R.A. Andersen, Encoding of intention and spatial location in the posterior parietal cortex, *Cereb. Cortex* 5 (1995) 456–469.
- [3] J. Artieda, M.A. Pastor, F. Lacruz, J.A. Obeso, Temporal discrimination is abnormal in Parkinson's disease, *Brain* 115 (1992) 199–210.
- [4] R.B. Banati, G.W. Goerres, C. Tjoa, J.P. Aggleton, P. Grasby, The functional anatomy of visual–tactile integration in man: a study using positron emission tomography, *Neuropsychologia* 38 (2000) 115–124.
- [5] K. Baynes, M.G. Funnell, C.A. Fowler, Hemispheric contributions to the integration of visual and auditory information in speech perception, *Percept. Psychophys.* 55 (1994) 633–641.
- [6] C. Bruce, R. Desimone, C.G. Gross, Visual properties of neurons in a polysensory area in superior temporal sulcus of the macaque, *J. Neurophysiol.* 46 (1981) 369–384.
- [7] K.O. Bushara, J. Grafman, M. Hallett, Neural correlates of auditory–visual stimulus onset asynchrony detection, *J. Neurosci.* 21 (2001) 300–304.
- [8] R. Cabeza, L. Nyberg, Imaging cognition II: an empirical review of 275 PET and fMRI studies, *J. Cogn. Neurosci.* 12 (2000) 1–47.
- [9] G. Calvert, P. Hansen, S. Iversen, M. Brammer, Crossmodal integration of non-speech audio–visual stimuli, in: *Human Brain Mapping*, Academic Press, San Antonio, TX, 2000.
- [10] G. Calvert, D. Lloyd, M. Brammer, How does the brain combine information from the different senses?, in: *Multisensory Research Conference*, Tarrytown Hilton, New York, 2000.
- [11] G.A. Calvert, E.T. Bullmore, M.J. Brammer, R. Campbell, S.C.R. Williams, P.K. McGuire, P.W.R. Woodruff, S.D. Iversen, A.S. David, Activation of auditory cortex during silent lipreading, *Nature* 276 (1997) 593–596.
- [12] G.A. Calvert, M.J. Brammer, S.D. Iversen, Crossmodal identification, *Trends Cogn. Sci.* 2 (1998) 247–253.
- [13] G.A. Calvert, R. Campbell, M.J. Brammer, Evidence from functional magnetic resonance imaging of crossmodal binding in the human heteromodal cortex, *Curr. Biol.* 10 (2000) 649–657.
- [14] G.A. Calvert, M.J. Brammer, E.T. Bullmore, R. Campbell, S.D. Iversen, A.S. David, Response amplification in sensory-specific cortices during crossmodal binding, *Neuroreport* 10 (1999) 2619–2623.
- [15] C. Carter, T.S. Braver, D.M. Barch, M.M. Botvinich, D. Noll, J.D. Cohen, Anterior cingulate cortex, error detection, and the online monitoring of performance, *Science* 280 (1998) 747–749.
- [16] J.D. Cohen, B. MacWhinney, M. Flatt, J. Provost, Psyscope: a new graphic interactive environment for designing psychology experiments, *Behav. Res. Methods Instrum. Computers* 25 (1993) 257–271.
- [17] C.G. Cusick, Superior temporal polysensory region in monkeys, in: K.S. Rockland, J.H. Kaas, A. Peters (Eds.), *Cerebral Cortex: Extrastriate Cortex in Primates*, Plenum, New York, 1997, pp. 435–468.
- [18] H.M. Duvernoy, in: *The Human Brain. Surface, Blood Supply, and Three-Dimensional Sectional Anatomy*, Vol. II, Springer-Verlag, Wien, 1999.
- [19] D. Embick, A. Marantz, Y. Miyahisa, W. O'Neil, K.L. Sakai, A syntactic specialization for Broca's area, *Proc. Natl. Acad. Sci. USA* 97 (2000) 6150–6154.
- [20] G. Ettliger, W. Wilson, Cross-modal performance: behavioural processes, phylogenetic considerations and neural mechanisms, *Behav. Brain Res.* 40 (1990) 169–192.
- [21] G. Ettliger, H.S. Garcha, Cross-modal recognition by the monkey: the effects of cortical removals, *Neuropsychologia* 18 (1980) 685–762.
- [22] J.A. Fiez, E.A. Raife, D.A. Balota, J.P. Schwarz, M.E. Raichle, S.E. Petersen, A positron emission tomography study of the short-term maintenance of verbal information, *J. Neurosci.* 16 (1996) 808–882.
- [23] N. Hadjikhani, P.E. Roland, Cross-modal transfer of information between the tactile and visual representations in the human brain: a positron emission tomography study, *J. Neurosci.* 18 (1998) 1072–1084.
- [24] G. Hickok, D. Poeppel, Towards a functional neuroanatomy of speech perception, *Trends Cogn. Sci.* 4 (2000) 131–138.
- [25] W. Horster, A. Rivers, B. Schuster, G. Ettliger, W. Skreczek, W. Hesse, The neural structures involved in cross-modal recognition and tactile discrimination performance: an investigation using 2DG, *Behav. Brain Res.* 33 (1989) 209–277.
- [26] T. Hosokawa, R. Nakamura, N. Shibuya, Monotic and dichotic fusion thresholds in patients with unilateral subcortical lesions, *Neuropsychologia* 19 (1981) 241–247.
- [27] M. Iacoboni, R.P. Woods, M. Brass, H. Beddering, J.C. Mazziotta, G. Rizzolatti, Cortical mechanisms of human imitation, *Science* 286 (1999) 2526–2528.
- [28] S. Kalaska, D.J. Crammond, Deciding not to GO: neuronal correlates of response selection in a GO/NOGO task in primate premotor and parietal cortex, *Cereb. Cortex* 5 (1995) 410–428.
- [29] J.R. Lackner, H.-L. Teuber, Alterations in auditory fusion thresholds after cerebral injury in man, *Neuropsychologia* 11 (1973) 409–415.
- [30] J.W. Lewis, M.S. Beauchamp, E.A. DeYoe, A comparison of visual and auditory motion processing in human cerebral cortex, *Cereb. Cortex* 10 (2000) 873–888.
- [31] A.M. Liberman, I.G. Mattingly, The motor theory of speech perception revised, *Cognition* 21 (1985) 1–36.

- [32] M. McGrath, Q. Summerfield, Intermodal timing relations and audio–visual speech recognition by normal-hearing adults, *J. Acoust. Soc. Am.* 77 (1983) 678–685.
- [33] H. McGurk, J.W. MacDonald, Hearing lips and seeing voices, *Nature* 264 (1976) 746–748.
- [34] M.A. Meredith, J.W. Nemitz, B.E. Stein, Determinants of multisensory integration in superior colliculus neurons: I. Temporal factors, *J. Neurosci.* 10 (1987) 3215–3229.
- [35] N. Nishitani, R. Hari, Temporal dynamics of cortical representation for action, *Proc. Natl. Acad. Sci. USA* 97 (2000) 913–918.
- [36] C.R. Olson, A.M. Graybiel, Sensory maps in the claustrum of the cat, *Nature* 288 (1980) 479–481.
- [37] M.W. Oram, D.I. Perrett, Responses of anterior superior temporal polysensory (STPa) neurons to ‘biological motion’ stimuli, *J. Cogn. Neurosci.* 6 (1994) 99–116.
- [38] J.V. Pardo, P.J. Pardo, K.W. Janer, M.E. Raichle, The anterior cingulate cortex mediates processing selection in the Stroop attentional conflict paradigm, *Proc. Natl. Acad. Sci. USA* 87 (1990) 256–259.
- [39] D.I. Perrett, P.A.J. Smith, A.J. Mistlin, A.J. Chitty, A.S. Head, D.D. Potter, R. Broennimann, A.D. Milner, M.A. Jeeves, Visual analysis of body movements by neurones in the temporal cortex of the macaque monkey: a preliminary report, *Behav. Brain Res.* 16 (1985) 153–170.
- [40] D. Poeppel, A critical review of PET studies of phonological processing, *Brain Language* 55 (1996) 317–351.
- [41] B.E. Stein, M.T. Wallace, Comparisons of crossmodality integration in midbrain and cortex, *Prog. Brain Res.* 112 (1996) 289–299.
- [42] Q. Summerfield, Lipreading and audio–visual speech perception, in: V. Bruce, A. Cowey et al. (Eds.), *Processing the Facial Image*, Clarendon Press/Oxford Press, Oxford, UK, 1992, pp. 71–78.
- [43] J. Talairach, P. Tournoux, *Co-planar Stereotaxic Atlas of the Human Brain*, Thieme Medical, New York, 1988.
- [44] R.F. Thompson, J.A. Shaw, Behavioral correlates of evoked activity recorded from association areas of the cerebral cortex, *J. Comp. Physiol. Psychol.* 60 (1965) 329–339.