Chapter 5

Audiovisual speech binding: perception and brain activity as a function of synchronicity

S. Tomaskovic, J.E.C. Wiersinga-Post, L.M. Slabu, R. Renken, and H. Duifhuis
Abstract

In the current study functional magnetic resonance imaging was used to investigate cross-modal integration of audio-visual (AV) speech perception by using the McGurk effect. In the McGurk effect, the auditory perception is changed due to AV integration. In our experiment fifteen subjects were asked to identify vowel-consonant-vowel syllables (/aba/, /ada/, /apa/, /ata/) presented auditory, visually and audiovisually. In order to manipulate the degree of audio-visual integration, we used stimuli with different time delays (-510, -110, +50, +210, +410 ms) between audio and visual information. We recorded subjects’ responses, and fewer responses on /aba/ and /apa/ stimuli indicated the stronger AV integration. On the basis that the strength of the McGurk effect reflects the strength of AV integration, we expected to find a correlation in brain activation and AV synchronicity. We observed the inferior parietal lobule, the supramarginal gyrus, and the superior frontal gyrus to be involved in AV integration.

5.1. Introduction

To perceive objects and events in the environment organisms are equipped with different senses, each specialized in detecting certain physical aspects of the outside world. Often, these objects or events emit signals that can simultaneously be detected by different senses. Speech, for example, can be perceived by listening to the speech sound, but also by watching the articulating movements of the mouth. Combining auditory and visual information improves speech understanding, especially when the acoustical signal is contaminated with noise [Sumby and Pollack, 1954]. How the brain manages to combine information that enters via different sensory channels and how the brain forms a unified percept of the outside world is a hardly known.

To date, many brain areas have been identified as possibly involved in audiovisual integration. These include high order cortical association areas, parts of the classically unimodal auditory and visual cortices, and subcortical areas [Calvert, 2001, Calvert et al., 2001, Laurienti et al., 2005, Miller and D’Esposito, 2005]. Several brain areas such as superior temporal sulcus, the intraparietal sulcus, the inferior frontal gyrus, the insula, claustrum, superior colliculus, visual motion cortex (MT/V5) and the primary auditory cortex are repeatedly mentioned in different studies. However, many discrepancies between different studies still remain. Furthermore, the functional relationship between brain activity and the perception of audiovisual stimuli remains largely unclear. Combining brain imaging with psychophysical measurements will provide further insight in this functional relationship.

The opportunity of modern brain imaging techniques to relate brain activation patterns to their perceptual correlates has not yet been fully exploited in
studies concerning audiovisual integration or binding, despite the fact that this is relatively easy due to the negligible invasiveness of these techniques and the related possibility to use human subjects. Only a few studies use behavioral responses that provide evidence for cross-modal binding [Bushara et al., 2003, Jones and Callan, 2003, Miller and D'Esposito, 2005].

Bushara et al. (2003) used an event related fMRI paradigm in which subjects watched two approaching moving bars and hear a collision sound. Identical auditory and visual stimuli could be perceived as bouncing bars, indicating that the auditory and visual information was bound together, or as two bars that passed, indicating that the audiovisual information was processed as unbound unimodal event. The brain activation patterns suggested that cross-modal binding was associated with higher activity in multimodal areas, such as the insula/frontal operculum, dorsolateral and medial prefrontal cortex, posterior parietal cortex, posterior thalamus, superior colliculus and posterior cerebellar vermis, but with lower activity in predominantly unimodal areas.

Miller and D’Esposito (2005) used an event related fMRI paradigm in combination with a simultaneity judgment task. Results showed a positive correlation between the number of stimuli that were perceived as being synchronous or fused and activity in several brain area’s including Heschl’s gyrus, the superior temporal sulcus, the middle intraparietal sulcus and the inferior frontal gyrus.

Feedback on whether or not audiovisual binding has taken place can also be derived by using the McGurk effect. To elicit this effect, an auditory speech stimulus, e.g. the vowel-consonant-vowel (VCV) /aba/, is presented together with incongruent articulating movements of a mouth producing a VCV sound with a non labial consonant e.g. /aga/. Fusion of this incongruent audiovisual information easily results in the perception of a similar sounding, but non labial consonant e.g. /ada/ [McGurk and MacDonald, 1976]. Jones and Callan (2003) use the McGurk effect as an indication for audiovisual binding. The speech sound /aba/ was paired with the video of a speaker pronouncing either /aba/, the congruent stimulus, or /ava/, the incongruent stimulus. The degree of audiovisual integration was manipulated by presenting these stimuli synchronous or 400 ms desynchronized. Subjects reported whether they heard the sound /aba/ or something else. At the group level, a regression analysis was used to relate the mean strength of the McGurk effect, i.e. the mean number of non-/aba/ responses, to levels of brain activation. The results showed a reversed correlation between the strength of the McGurk effect and activity in the left occipitotemporal junction.

In this study we also use the McGurk effect as an indicator of audiovisual binding. The VCV syllables /aba/ and /apa/ were dubbed on the movies of a mouth pronouncing the VCV’s containing the non labials, /aga/ and /aka/. Because the consonants /b/ and /p/ are easily confused with the consonants /d/ and /t/ respectively [Miller and Nicely, 1955] and because these are very similar to the visemes /g/ and /k/, subjects sensitive to the McGurk effect will perceive /ada/ and /ata/ when listening to these incongruent stimuli.
To select subjects that were sensitive to the McGurk effect, extensive psychophysical experiments were performed prior to the fMRI experiments. The synchronization of the audiovisual stimuli was varied between -620 and +420 ms, which enabled us to select stimuli for which the strength of the McGurk effect was maximal, minimal and intermediate. These stimuli were used in the fMRI experiments.

To ensure the audibility of the consonants in the VCV stimuli, we used a sparse sampling paradigm to reduce the influence of scanner noise [Hall et al., 1999].

Our results were partially in agreement with previous studies, finding posterior parietal and superior frontal areas involved in audio-visual integration. However, we did not find claustrum nor the superior temporal sulcus active, and those areas were also reported as areas involved in AV integration [Calvert, 2001].

5.2. Materials and methods

5.2.1. Subjects

Twenty healthy volunteers (aged 22 - 45, 7 females) with normal hearing and normal or corrected to normal vision participated in the fMRI study. None of them reported neurological or language impairments. One of these subjects was left-handed. Four out of the initial pool of 20 subjects (2 females) were excluded from the fMRI data analysis because of the absence of the McGurk effect as revealed in the psychophysical study. One subject did not complete the scanning experiment. In total we analyzed fifteen subjects that showed sensitivity to the McGurk effect. Prior to participating, subjects signed an informed consent. This study has been approved by the local Medical Ethical Committee, Groningen.

5.2.2. Stimuli

Movies were made with a digital video camera (Panasonic 3 CCD 24x digital zoom) and external microphone (Sennheiser) in which a female Dutch speaker uttered the vowel-consonant-vowel (VCV) syllables /aba/, /aga/, /apa/ and /aka/. The video frame rate was 25 Hz and the sounds were digitized at a sampling rate of 48 kHz and a 16-bit resolution. The movies were edited in Adobe Premiere. Two incongruent stimuli were made in which the audio of the /aba/ and /apa/ utterances were dubbed on the videos of the speaker saying /aga/ and /aka/, respectively. Stimuli with different synchronization were made ranging from -620 (audio leads video) to +420 ms (video leads audio). The duration of the acoustic VCV syllables was around 730 ms. The total length of the audiovisual movies was
2 s with the auditory utterance temporally centered. Only the mouth was included in the video frame. To avoid onset/offset information due to the opening and closing of the mouth, the movies were edited such that they started and ended with an open mouth. Thus, only the consonant within the VCV stimulus provided information about the synchronization of the audiovisual stimulus.

In the psychophysical experiment audiovisual (AV) stimuli were used with fifteen audiovisual delays ranging from -630 to +410 ms (-630, -510, -390, -270, -190, -110, -70, -30, +10, +50, +90, +130, +210, +290 and +410 ms). The negative delay implied that audio led video and for a positive delay, video led audio. Based on results of the psychophysical experiment, stimuli were selected for the fMRI experiment that resulted in a minimal (-510 and +410 ms), intermediate (-110 and +210 ms) and maximal (+50 ms) strength of the McGurk effect. Further, audio-only (A) stimuli, in which the speech sounds /aba/ or /apa/ were combined with a blank screen with a fixation cross, and visual-only (V) stimuli, in which movies of the mouth pronouncing /aga/ or /aka/ were shown, were used. The baseline stimulus was a fixation cross on the screen.

In the psychophysical as well as the fMRI experiment, all stimuli were presented against a continuously present pink noise background of 40 dB(A). This relatively high pink noise level was used to mask environmental noise present in the fMRI environment. The signal to noise ratio of the auditory stimuli was 20 dB.

### 5.2.3. Psychophysics

Prior to the fMRI experiments, a purely psychophysical test was performed. Subjects were seated in a sound insulated booth (Industrial Acoustic Company). Auditory stimuli were presented via headphones (Telephonics TDH-49P) and visual stimuli were presented on a computer screen at a distance of approximately 60 cm. The stimuli were presented seven times, resulting in a total of 210 stimulus presentations. Stimuli were presented in pseudo random order. In a four-alternative-forced-choice task, subjects had to indicate whether they heard the sound /aba/, /ada/, /apa/ or /ata/ by pressing corresponding keys on a keyboard with their left or right index or middle finger. Each subject repeated the experiment 3 times.

### 5.2.4. fMRI experimental design

Functional MR scans were acquired using a 3T scanner (Philips Intera). Prior to the functional images, a T1-weighted anatomical scan of the entire brain was obtained (160 slices; FOV = 256 mm; matrix 256 x 256; transverse slice orientation; TR = 25 ms; TE = 4.6 ms; flip angle 30°). Functional MR images were acquired using echo planar imaging (46 slices; slice thickness 3.5 mm; field of view (FOV) 224 mm; scanning matrix 64 x 64; transverse slice orientation;
repetition time (TR) = 20 s; echo time (TE) = 35 ms; flip angle 90°). Eight different stimulus conditions were used: 1 to 5: AV with delays of -510, -110, 50, 210 and 410 ms, respectively; 6: A-only; 7: V-only and 8: baseline. Each experiment consisted of two functional series in which seventy-four volumes were scanned (18 volumes of the baseline condition and 56 volumes, in which the seven remaining stimulus conditions were presented 8 times). To avoid interference with scanner noise during acoustic stimulus presentation, we applied the sparse sampling method [Hall et al., 1999] with a relatively long repetition time of 20 s. Between two acquisitions, 4 VCV stimuli of one stimulus condition were presented.

Stimuli were presented in pseudo random order, while care was taken to balance the number of /aba/+aga/ vs. /apa/+aka/ stimuli. The total experiment, including one anatomical scan and two functional series, lasted approximately 50 minutes. As in the psychophysical test, the subjects’ task was to indicate which consonant they heard, /b/, /d/, /p/ or /t/, by pressing a button on a response box with their left or right index or middle finger. Subjects had to respond within 1 s.

The stimulus presentation was followed by a window showing four buttons (at the same position as the mouth) corresponding to the possible answers: /aba/, /ada/, /apa/ and /ata/. Visual stimuli were projected onto a screen located at the base of the scanner via a LCD projector (Barco G300j). Subjects viewed the movies through a mirror angled above the subjects’ head in the scanner. Auditory stimuli were presented through headphones of the sound system of the scanner. Subjects wore earplugs and mufflers, which reduced the scanner noise by approximately 20 dB.

5.2.5. Data analysis

SPM 99 software (Welcome Department of Cognitive Neurology, London, UK; http://www.fil.ion.ucl.ac.uk/spm/spm99.html) was used to process and analyze the data. The EPI images were realigned to the first volume in the run. These images were coregistered with the subjects’ own anatomical image. Next they were normalized to the standard brain (MNI) and spatially smoothed using a Gaussian kernel with a full width-half-maximum of 10 mm.

MarsBaR toolbox (http://marsbar.sourceforge.net) was used to extract regions of interest (ROI) and to plot estimates of block related responses.

Two step procedure in analysis was performed: first and second level analysis. On the first level analysis we applied block design for each individual. Analysis was done using a simple t-test between MR signal collected in the ‘on’ conditions versus the ‘off’ conditions. We used sparse sampling, therefore we did not convolved it with HRF.

In order to correlate subjects’ responses with brain activations a second level analysis was applied, where estimated $\beta$ values from the first level analysis were used. Average percentage of subjects’ responses per condition were calculated and then used in the further analysis. Two different second level analyses were performed. Firstly, averaged individual responses on the McGurk
stimuli were concatenated into one regressor. Next, the regressor was scaled 1, i.e., the mean value was set to 0, and standard deviation to 1. In second analysis, average percentage of subjects’ responses were scaled, and the scaled responses were concatenated into a regressor. Furthermore, what differs from previous analysis is that off-set of the BOLD responses were also scaled.

5.3. Results

5.3.1. Behavioral performance

Figure 5.1 shows the behavioral data of individual subjects, for both, data recorded during the scanning, and the psychophysical data measured prior to the fMRI experiment (from the acoustical booth). There is a common trend in subjects’ behavior, viz. the strongest McGurk effect is when visual leads the audio information for approximately +50 ms. For different time delays between audio and visual information, different strength of the McGurk effect is present. The percentage of responding to the McGurk stimuli varies among subjects. Range for maximum strength of the McGurk effect, for psychophysical data from the acoustical booth is between 45% and 100%, while for psychophysical data form the scanner is between 23% and 94%.

5.3.2. Unimodal processing

Figure 5.2 presents a map of the significant activation areas for unimodal visual stimulation. Bilateral activations were found in the frontal lobe (indicated by 1) - more specifically Broca’s area, parietal lobe (indicated by 2), occipital lobe (indicated by 3) - MT/V5 area, and cerebellum (indicated by 4). Not significant activation was found in the primary or secondary auditory cortices. Figure 5.3 shows the related map for unimodal auditory stimulations. In addition to the four areas mentioned above, now also activation is observed in the auditory cortices, in the temporal lobe (indicated by 5).
Figure 5.1. Individual behavioral data of 15 subjects from the fMRI experiment. On the x-axes are 5 AV stimuli, and on the y-axes are percentage of subjects’ responses to the McGurk stimuli. Grey circles correspond to psychophysical data from the acoustical booth, and black circles to subjects’ responses from the fMRI setup.
5.3.3. Cross-modal processing

The basic SPM analysis was preformed and direct comparison of the condition for the strongest McGurk effect (AV50) with the condition for the weakest McGurk effect (AV−510 or AV410) does not show any significant...
differences in the activation patterns. The same result was revealed for the two intermediate stimuli (AV110 and AV210) also contrasted with the strongest McGurk effect.

In the further analysis we contrasted the strongest AV50 condition (p < 0.005, uncorrected) to A (audio only) and masked by [AV50 vs. V] inclusively, so that the comparison is made using only voxels that reached significance (p < 0.005, uncorrected) in the [AV50 vs. V] condition. Still, we did not find significant activity.

Next, contrast that was used: [A vs. AV50] inclusively masked by [V vs. AV50], for both conditions p < 0.005 uncorrected. Significant increase was found in the right supramarginal gyrus (SMG) (fig. 5.4A), right inferior parietal lobule (IPL) (fig. 5.4B), and right superior frontal gyrus (SFG) (fig. 5.4C). Plots of strength of BOLD response for those three ROIs at: (62, -44, 28), (48, -58, 42), and (12, 20, 58) are shown on figure 5.4A, 5.4B, and 5.4C respectively (lower panel). The x-axis represents stimuli (different synchronization), and the y-axes represents the size of the effects (only relative ratios between conditions are presented). It is clear that BOLD response depends on the amount of synchronization.

Figure 5.4. Results of [A vs. AV50] inclusively masked by [V vs. AV50] indicating significant increase in A or V conditions relative to AV condition (p < 0.005 uncorrected). Upper panel shows the mean anatomical image and corresponding activation and lower panel represents ROI analysis showing percent signal change on the y-axes (indications 1-7 are stimuli: AV−510, AV−110, AV50, AV210, AV410, A, and V, respectively). (A) right supramarginal gyrus (62, -44, 28), (B) right inferior parietal lobule (48, -58, 42), and (C) right superior frontal gyrus (12, 20, 58).
5.3.4. Correlation of behavioral responses and BOLD signal

Figure 5.5B shows the SPM99 - $\beta$ values (signal changes for the stimulus conditions, with respect to the baseline) changes to 5 different stimuli, for the activated area with the maximum at (-42, -46, 26), indicated on figure 5.5A. Contrary to the previous analysis, this activation was found in the left hemisphere. It is clearly shown (in fig. 5.5B), that there is a common trend (of BOLD response)
present among the subjects, and it follows the trend already shown for psychophysical data. A similar analysis with normalized subjects’ responses (i.e. responses were scaled to minimize the amplitude differences between responses) shows a significant activation in the opposite hemisphere, at (62, -46, 26) (fig. 5.6A). The scaling factor that was used aimed to reduce the general variability in the subjects’ responses. Figure 5.6B shows again common trend (per subject) of BOLD response following the trend of psychophysical data.

Figure 5.6. (A) Mean anatomical image for activation of all 5 AV stimuli correlated to the behavioral responses (from the fMRI setup). The right supramarginal gyrus was found active (62, -46, 26), for p < 0.005. (B) β values plotted against 5 different stimuli for the same area as in A). Subjects’ responses were normalized.
5.4. Discussion

5.4.1. Behavioral performance

Recording the subjects’ responses from the scanner gives a possibility to investigate if our subjects were integrating audio and visual information or not. The subjects’ behavioral responses from the scanner are comparable to the psychophysical responses, on which we based selection of stimuli and subjects. Because of the scanner noise we used the sparse sampling method for data acquisition and the same stimuli type were presented within one block, while in the psychophysical experiment stimuli were presented randomly. That might cause some of the differences between those two data sets.

Figure 5.1 shows that all subjects behaved similarly, although subjects 9 and 11 do not reach the 50% score but still a different strength of the McGurk effect due to synchronization of AV stimuli is present. Furthermore, for the weakest McGurk effect subjects’ responses never drop to 0%. That was not surprising as there was always audio and visual information present, thus, it might suggest that there is always some audio-visual interaction.

The maximum strength of the McGurk effect was for approximately 50 ms where visual information lead the audio. Similar result was reported by Munhall et al. (1996) where the strongest McGurk effect was shown for 60 ms where audio lagged the visual information. It seems that the perfect synchronization of visual and auditory speech information is not necessary for audio-visual integration [McGrath and Summerfield, 1985, Munhall et al., 1996, Jones and Munhall, 1997, Miller and D’Esposito, 2005].

5.4.2. Unimodal processing

The MT area seems to be a mainly visually specific area involved in the visual lip-reading task. This result was expected, as MT area has been implicated in visual motion processing (Zeki, 1993). Although Broca’s area in the frontal cortex is often activated for the auditory stimuli it is active for the visual stimuli as well. Thus, this area is not an auditory only specific area. Furthermore, strong left lateralization can be seen in temporal lobe for unimodal auditory stimuli. Zatorre (2001) reported that tonal processing recruits mechanisms in the right auditory cortex, while speech like stimuli are processed in the right auditory cortex.

The main difference between V only and A only conditions was the activations in temporal lobe, primarily in auditory cortex for A only condition (fig. 5.3), while for V only (fig. 5.2) the same areas were not found to be active. For unimodal visual stimuli Calvert et al. (1997) reported primary auditory activation.
Later, this primary auditory activation was not replicated in the number of studies [Campbell et al., 2001, Bernstein et al., 2002] and neither in our study.

5.4.3. Cross-modal processing

By altering time delays between audio and visual information we produced different AV situations, i.e., different strengths of AV integration. Figure 5.1 shows that even for the extreme time delays between audio and visual information, subjects still responded as combining those two cues. Based on those data, we might suspect that there is always AV interaction present, but in different degrees. That might explain why direct comparison between the strongest McGurk effect (AV50) and weakest strength of it (AV−510 or AV410) does not give significant activation. It seems, for our stimuli, the brain always integrated audio and visual information present and we were not able to find activations due to different stimuli. Thus, for all 5 stimuli the same areas were active during processing: areas involved in audio processing, areas involved in visual processing, but also areas involved in AV processing and integration.

The principle finding in this study is the inverse correlation of three areas we found active (IPL, SMG, and SFG) with synchronization of stimuli. It seems that for the stronger McGurk effect, there is less activation, thus, possibly our brain works less when processing AV stimuli that produce strong McGurk effect, while for more desynchronized stimuli (i.e. weaker McGurk effect), the delay between stimuli is obvious and more confusing and that produces more activation. IPL and SMG are generally thought to be seen as a higher order association areas and they are considered to contain cells that respond to both auditory and visual stimuli, previously identified in animals [Andersen, 1995]. IPL and SMG receive multimodal inputs and have already been identified as areas involved in sensory, motor, and attentional functions [Jones and Callan, 2003]. The SMG plays important role in the perception and production of phonemes [Caplan et al., 1995, Ojemann et al., 1989], in a categorization of phonemes [Zatorre et al., 1996] and it contributes to a short-term memory task for speech related tasks [Paulesu et al., 1993]. Miller and D’Esposito (2005) suggested that inferior parietal areas serve as a sensory-motor interface; those areas are high-level association cortices with massive cross-modal convergence. The finding that IPL and SMG are involved in AV processing was not unexpected.

A secondary finding in this study is activation in the superior frontal lobe, for the inclusive masking condition mentioned above. The involvement of the frontal cortex is usually related to reasoning, planning, and problem solving. In cross-modal processing, function of frontal cortex is still less well understood, but there is some evidence of its involvement in integration of arbitrary cross-modal cues or associations [Calvert, 2001].

Although, comparison between different stimuli does not demonstrate any activation due to different strength of the McGurk effect, investigating ROI analysis for SMG, IPL and SFG activations (lower panel fig. 5.4A, 5.4B, and 5.4C)
suggested there is common trend in a percent signal change for different stimuli. And again the strongest percent signal change is for AV50, thus for the stimuli where the McGurk effect is the strongest.

We expected to find activations in superior temporal sulcus (STS), which was commonly reported as an area involved in AV integration [Miller and D’Esposito, 2005, Sekiyama et al., 2003, Calvert et al., 2000]. Electrophysiological studies have shown that the STS contains cells that respond to stimulations in more than one sensory modality [Desimone and Gross, 1979, Sekiyama et al., 2003]. However, Jones and Callan (2003) did not find STS activations, which is in agreement with our study. They compared congruent and incongruent conditions and there were no evidence of activations in the region of STS. Conversely, when they examined t-maps of synchronous and asynchronous incongruent conditions; more extensive activity in the temporal lobe was observed for conditions where audio and visual signal were out of phase [Jones and Callan, 2003]. It is in accordance with our result, comparing t-map AV50 vs. baseline with t-map AV−410 vs. baseline, the most extensive temporal activity was found for the largest asynchronous stimuli (AV−410) when audio leads the visual stimuli, although direct comparison between stimuli does not show any significant differences in the STS.

5.4.4. Correlation of behavioral responses and BOLD signal

The main advantage of combining the subjects’ responses with fMRI analysis is detection of brain areas that are correlated to the subjects’ behavior. Although still unclear why, analysis with non normalized subjects’ responses included, gave us an activation in the left SMG (fig. 5.5A), thus activation at different location than in previous analysis (fig. 5.4A). Once again if we examine figure 5.1 it is clear that the common behavioral trend for synchronization does not differ too much among subjects but the strength of subjects’ responses does. That might be the reason of this different location of activation. Our explanation furthermore can be confirmed with the result of the next analysis; subjects’ responses were scaled, thus the amplitude differences between those responses were excluded. Figure 5.6A shows this result and it coincides with the result shown on figure 5.4A. Thus, from three different analyses, two results show the same areas of activation (fig. 5.4A and fig. 5.6A) where the amplitude differences of subjects’ responses were excluded.

A different area of activation was presented in figure 5.5A, where the intersubject variabilities for amplitude were included. That is why we speculate that the result depends on inter-subject variabilities.

Additionally, we plotted β values against different stimuli (fig. 5.5B, and fig. 5.6B) per subject, and for both left SMG (-42, -46, 26), and right SMG (62, -46, 26), to compare how the BOLD signal of subjects differentiate. Although, it is obvious that there are differences between subjects, there is noticeable common trend present among subjects, which is comparable between all of them. Moreover,
the graphs on figure 5.5B, and figure 5.6B are comparable with behavioral graphs (from acoustical booth and fMRI setup), and again common trend is present. These graphs confirm our assumption that there is correlation between subjects’ responses and the BOLD signal.

This is still the work in process. Although we have some indications, for more solid conclusions and understanding extra analysis have to be done. The first step is: the single-subject analysis that should indicate what areas are active due to AV integration (per subject), and areas of correlated subjects’ responses and BOLD signal. Analysis preformed until now gave us different results, and single subject analysis maybe have a power to explain what is still not understandable.

5.5. Conclusions

We set up a series of psychophysical and fMRI experiments to investigate how the strength of the AV binding depends on AV synchronicity. Furthermore, we examined if there is correlation between the strength of AV integration and brain activations. In this paper, behavioral responses from the fMRI coincide with our psychophysical results, measured prior to the fMRI experiment. As previous studies already observed, the strongest McGurk effect was found when visual information leads the audio, in our experiment that was for +50 ms.

We identified three areas that are involved in AV integration, activations in the SMG, IPL, and SFG. The posterior parietal cortex is a polymodal region and our results suggest it. Although less understood, there is an indication that superior frontal gyrus is involved in integration of associative cross-modal cues.

5.6. References


