Audiovisual processing of Chinese speech sound - character pairs - a MEG study

Katja Koskialho

Pro gradu -tutkielma

Psykologian laitos

Jyväskylän Yliopisto

Elokuu 2016

JYVÄSKYLÄN YLIOPISTO
Psykologian laitos
KOSKIALHO, KATJA: Audiovisual processing of Chinese speech sound - character pairs - a MEG study
Pro gradu -tutkielma, 23 s, x liites.
Ohjaaja: Jarmo Hämäläinen
Psykologia
Elokuu 2016

Tämän tutkimuksen tarkoituksena oli selvittää audiovisuaalisen integraation aivoperustaa. Tutkimuksen toteutuksessa käytettiin visuaalisena ärsykkeenä kiinankielisiä kirjoitusmerkkejä ja auditorisena ärsykkeenä kiiniankielisiä puheäänteitä, joiden vaikutusta aivoihin tarkasteltiin joko esittämällä niitä unimodaalisesti tai bimodaalisesti. Muodostettiin kaksi tutkittavien ryhmää: suomenkieliset, jotka eivät osanneet kiinaa sekä natiivit kiinan puhujat, jotka ymmärsivät myös merkkien ja puheäänteiden merkityksen. Näin päästiin tarkastelemaan myös sitä, onko sanojen merkityksen tuntemisella merkitystä äänteiden ja merkkien prosessointiin.

Aivojen toimintaa mitattiin 306-kanavaisella aivomagneettikäyrää mittaavalla magnetoenkefalografilla, MEG:llä. Aivoaktiivisuuden paikantumista tutkittiin BESA 6.1. -ohjelmistolla, Classical LORETA Analysis Recursively Applied (CLARA) - lähteenpaikannuksella. Suomalaisten ja kiinalaisten aivoaktiivisuutta vertailtiin eri tilanteiden välillä. Tulokset osoittivat, että suomalaisten visuaalinen aivokuori aktivoitui voimakkaammin kuin kiinalaisten, kun visuaalisia ärsykkeitä esitettiin ilman ääntä (V). Kun stimuluspareja (AV) esitettiin kongruentisti eli yhdenmukaisesti, suomalaisten oikea ja vasen fusiforminen poimu aktivoitui voimakkaammin kuin kiinalaisten. Tuloksista voidaan päätellä, että audiovisuaalinen prosessointi aivoissa toimii monimutkaisesti, ja merkityksen ymmärtämisellä voi olla vaikutusta siihen, millä aivoalueella tai miten voimakkaasti aivot aktivoituvat.

Avainsanat: audiovisual integration, audiovisual processing, speech sounds, multisensory integration, fusiform gurys, posterior superior temporal sulcus

UNIVERSITY OF JYVÄSKYLÄ
Department of Psychology
KOSKIALHO, KATJA: Audiovisual processing of Chinese speech sound - character pairs - a MEG study
Master's thesis, 23 p., x pp.
Supervisor: Jarmo Hämäläinen
Psychology
August 2016

The aim of this thesis was to find out the neuronal basis of audiovisual integration. In the implementation of this study Chinese characters were used as a visual stimulus and for the auditory stimulus Chinese speech sounds were used, of which effects to the brain were observed by presenting the stimuli either unimodally or bimodally. Two groups of subjects were formed: Finnish subjects who did not know any Chinese and native Chinese subjects who also understood the meaning of the characters and speech sounds. This made it possible to also examine whether understanding the meaning of words made an impact on the processing of speech sounds and characters.

The brain activation was measured by 306-channel magnetic field-recording magnetoencephalogram, MEG. The localization of the brain activation was explored by BESA 6.1. software, Classical LORETA Analysis Recursively Applied (CLARA) - source localization. The brain activation of the Finnish and Chinese subjects were compared between different conditions. The results show that Finnish subjects' visual cortex activated more strongly than Chinese, when the visual stimulus was presented without the voice (V). When stimulus pairs (AV) were presented congruently or consistently, Finnish right and left fusiform gyrus was activated more powerfully than Chinese. A conclusion can be reached that the audiovisual processing in the brain functions in complex ways, and knowing the meaning of the shown stimuli can affect in which brain region or how powerfully the brain activates.

Key words: audiovisual integration, audiovisual processing, speech sounds, multisensory integration, fusiform gurys, posterior superior temporal sulcus

CONTENTS

INTRODUCTION

- Language in the brain
- Audiovisual integration and learning
- The recent research field of audiovisual integration
- Temporal features of the integration
- Superior temporal sulcus - behind everything?
- Magnetoencephalogram, MEG
- The aims and research questions

METHODS

- Subjects
- Stimuli and experimental design
- Equipment
- Data processing
- Statistics

RESULTS

DISCUSSION

- Activation in the cortex: Discovered differences between the groups and conditions
- The role of fusiform gyrus in the speech sound - letter processing
- Critique
- Further study
- Acknowledgements

INTRODUCTION

Language in the brain

Vision is probably humans' most acute sense. Human vision is a complex but widely studied and well understood system. Visual processing is performed largely on the occipital lobe, in the visual cortex. In addition to vision, sense of hearing and especially speech perception is also a well explored system. Auditory perception is based on changes in air pressure which is transformed to electric impulses and projected then to the primary auditory cortex, A1.

Basic seeing and hearing are parts of the coherent whole perception: both capabilities among other senses are needed in order to create a perception. Visual and auditory information must be combined together to perform several ordinary human functions, for example speaking or reading or other common hobbies. In everyday life, human brain receives input constantly from the environment in different kinds of sensory forms. While driving a car we see the traffic lights, hear the engine, observe other moving cars around us and spot the walkers on the side of the road, hear when someone honks and smell old car's exhaust smoke. Besides all that we sense the steering wheel in our hands and a pedal under the foot. These are not separate sensations: by integrating all the sensory information, a whole sensation of driving is created into a unified percept.

Another everyday life example is reading. When learning to read, the pupil must learn to combine a specific speech sound to the certain visual character, the letter. By combining these two sensations, *audiovisual integration* is needed. Both are needed to perform a successful reading, primary visual cortex and primary auditory cortex. but they alone are not enough. In addition to these areas, regions for speech sound processing are also important.

For language and speech, there are several identified anatomical language areas in the brain. The classic speech and language areas are Broca's and Wernicke's areas, located laterally at the left side of the brain, in the cortex. In Brodmann's mapping, Broca's area is found in the areas 45 and 44, as for Wernicke the area is 22. Anatomically speaking, these language processing areas are located between inferior frontal gyrus and the superior temporal gyrus. Underneath the lateral fissure there are more structures concerning language: the insula, Heschl's gyrus and some parts of the superior temporal gyrus. (Kolb & Whishaw, 2015.)

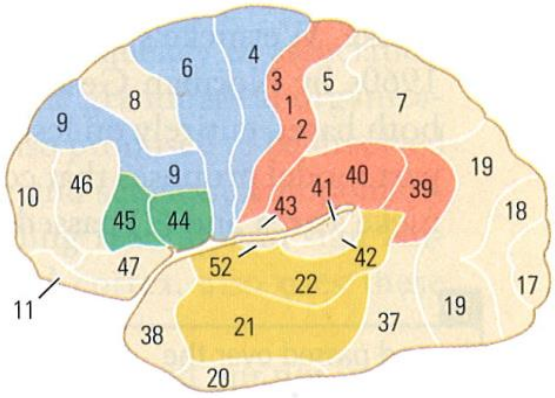

Figure 1. Brodmann's mapping. Broca's area in 45 and 44, Wernicke 22. (Kolb & Whishaw, 2015.)

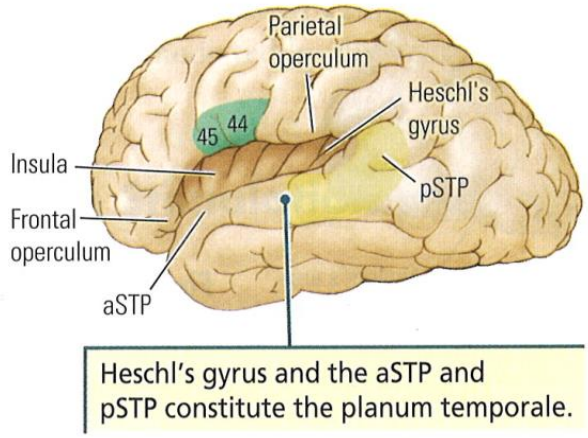

Figure 2. Other language relevant structures on laterally located in the brain. (Kolb & Whishaw, 2015.)

Traditionally speaking, Broca's area is usually considered as a speech production area and Wernicke's as speech comprehension area. Since the areas have both been found at the late 1800's, the information of these two has been gathered a lot more during the past century. Several subdivisions based on transmitter receptors and cytoarchitecture in the Broca's area have been defined (Amunts et al., 2010) and also other several human functions have been connected to the Broca's area: Brodmann's area 44 and 45 has also been linked to motor tasks (Iacoboni & Wilson, 2006)), integration of semantic information from speech and gestures (Willems et al., 2007) and other tasks demanding cognitive abilities (Herath et al. 2001, Novick et al., 2005).

The ability to integrate different sensory modalities is not as well understood as vision or auditory systems alone, and Broca's and Wernicke's areas are large anatomical regions

including several important but not well understood smaller structures. However, despite the complexity of the phenomenon, multisensory integration is still much studied. Because for the ordinary human behavior the multisensory integration is absolute necessity, it is a big question for the neuropsychological research field, how does this integration happen at the physiological level (Calvert, 2004).

Audiovisual integration and learning

In a brain level, sense of touch and visual and auditory systems are complex alone, and this is why integrating sensory systems has been difficult to study. However, sensory integration has been explored widely and from several different perspectives.

Reading is probably the most ordinary function in everyday life which demands audiovisual integration. What makes this more interesting, reading is still evolutionary speaking a young skill and property of the brain. In humans, learning how to separate speech sounds and learning how to speak begins already in utero (Goméz & Gerken, 2000). Although speech production and speech comprehension both have their own anatomical locations in the brain, reading combines computations from several other brain networks. To combine a speech sound to specific letter has been studied widely (e.g. Van Atteveldt et al., 2004) and especially when there appears problems in these speech sound representations (Froyen et al., 2011) and also anatomical features of injured brains when it comes to coordinating speech articulation (Dronkes, 1996). However, what exactly happens in the brain when a person is reading isn't still all clear despite all of the research.

The recent research field of audiovisual integration

When learning to combine a visual character to a specific speech sound the brain must integrate information coming from different channels. This multisensory integration has raised several questions among brain researches: Where does the auditory and visual integration happen in the

brain, is there a specific integration-location? Does the integration affect temporal qualities of processing, when compared to regular, unimodal stimulus perception? If there is two different sensory stimuli which needs to be integrated instead of one, is the procession time for those stimuli shorter or longer? Also, in order to integrate two different sensory modalities, both events must occur closely in time and space (Bolognini et al., 2005, Frassinetti et al., 2002).

Different brain regions have been suggested to be critical in multisensory integration. Many studies have shown that posterior superior temporal sulcus (pSTS) and middle temporal gyrus (MTG) are places that activate during integration (e.g. Beauchamp et al., 2004 and Gilbert et al., 2013). Beauchamp et al. (2004) studied audiovisual integration by functional magnetic resonance imaging, fMRI. Subjects were presented with either pictures or sounds of animals or tools, first as single (either picture or sound) and then within pairs. When shown in pairs, subjects decided whether the object was congruent or incongruent to the sound. In the results Beauchamp et al. reported that in both hemispheres, pSTS/MTG, dorsolateral prefrontal cortex (DLPFC) and ventral temporal cortex responded more strongly to AV blocks than to single auditory or visual blocks.

In Gilbert and colleagues' (2013) research participants had learnt 12 different stimulus pairs. Pairs were formed to audiovisual pair, audio-audio pair, visuo-visual pair and visual-audio pairs. After a learning period, subjects' performance were recorded during magnetoencaphalogram, MEG, and they were asked whether the shown pairs were correct or not. Their results suggest increased activity on the right pSTS area, when crossmodal pairs (AV or VA) were shown. This area also seemed to be responding to auditory stimuli.

Slightly similar research was carried out by Tanabe, Honda and Sadato (2005). In their research, participants were asked to identify AV or VV pairs by trial and error. After giving the answer, participants had feedback whether the response was correct or not. The fMRI signal in the superior temporal sulcus (STS) suggests that the area could be important to the creation of paired stimulus, regardless of the stimulus type. During the delay period between the presentation of crossmodal paired stimuli there was activation in pSTS area, and no activation during unimodal paired associates.

Calvert et al. (2000) also explored audiovisual integration, although their experimental settings differed a bit from the ones above. As the auditory stimuli, speech sounds were played, and as visual stimuli mouthed speech was shown to the participants. These stimuli were shown both separately and at the same time, and also the mouthed speech was either congruent or

incongruent to the heard speech. The only place where the audiovisual integration could be localized was ventral bank of the superior temporal sulcus (STS) in the left hemisphere.

When it comes to localization, (posterior) superior temporal sulcus and middle temporal gyrus both seems strongly linked to integration process according to previous research. But does the integration change somehow, when the stimuli particularly consist of letters and speech sounds? Hashimoto and Sakai (2004) has explored letter learning in adulthood. In their fMRI study, Hangul letters, which the participants were not familiar with, were played either with the speech sound or not. When played with speech sound, the left posterior inferior temporal gyrus (PITG) and the parieto-occipital cortex (PO) were activated, they report. They suggest that the left PITG and the left PO play a primary role when forming a link between a letter and speech sound, and these structures have a critical role in cortical plasticity when it comes to learning new letters. Hashimoto and Sakai haven't been the only ones who have had these kind of results. Many studies have reported results which connect the link between brain activation for letters in the left PITG-region (Paulesu et al., 2000, Mechelli et al., 2003).

Not only learning the letter - speech sound combinations have been studied but also the meanings of these letter - sound combinations. In their paper, Vandenberghe, Price, Wise, Josephs and Frackowiak (1996) report exploring semantic features of words and pictures by positron-emission tomography, PET. As a result they reported activation in the left STS, left anterior MTG and in left inferior frontal sulcus during the semantic tasks concerning words.

In summary it seems that (p)STS, MTG, DPFC, PITG and PO areas are most likely included when it comes to integrating visual and auditory stimulus, some regions especially when it comes to speech sounds and letters. In addition to these studies also other results have been reported. In her feature article, Calvert (2001) reports different brain imaging techniques and findings concerning crossmodal processing in the brain. In her conclusion the insula is suggested to detect the crossmodal coincidence and it may also be involved in crossmodal matching. Calvert also confirms the role of STS in her review.

Exploring audiovisual integration by speech sounds and letters is large and common section of this research field. In addition to Hashimoto and Sakai, letter - speech sound combinations have been studied by Raij et al. 2000, van Atteveldt et al., 2004, Hertrich et al. 2007 and Tuomainen et al. 2005. By using fMRI, van Atteveldt et al. (2004) studied the functional neuroanatomy of the integration of letters and speech sounds. They as well located the letter/speech sound integration to the superior temporal cortex, but also anterior temporal

regions were connected to the bimodal congruent letters. They also reported that subregions of early auditory cortex, which are involved in speech sound processing, were influenced by the congruently played letters and speech sounds: the region did not respond to a letter alone, only when presented together with a speech sound.

Temporal features of the integration

Besides the localization, Raij, Uutela, and Hari (2000) have also investigated the temporal features of audiovisual integration in their study of neural representations of audiovisual objects, speech sounds and corresponding letters. The processing of the letters and speech sounds started at 60-120 ms. Both auditory and visual brain activations converged at 225 ms after stimulus onset, and interacted then mostly on the right temporo-occipito-parietal junction at 280-345 ms, and both in the left superior temporal sulci (380-540 ms) and right superior temporal sulci (450-535 ms). In addition to temporal qualities of this study, also the activated regions are corresponding to the previous studies.

As introduced earlier, Gilbert et al. (2005) explored auditory and visual stimuli by crossmodal and unimodal designs. In addition to spatial features of their findings, also temporal qualities were studied. The timing differences in pSTS were observed between crossmodal and unimodal trials. The earliest significant difference between crossmodal and unimodal stimuli occurred in a time window from 250 to 750 ms, which suits to previous research (e.g. Raij et al. 2000).

Superior temporal sulcus - behind everything?

Despite the complexity of the phenomenon, audiovisual learning and integration has a great role in the neuropsychological research field. A lot of different results have been reported, but not all results are conflicting. In general, the previous literature suggests superior temporal sulcus, STS, and a middle temporal gyrus, MTG, to have a critical role for comprehension of spoken and written words. (e.g. Gilbert et al. 2013, Tanabe et al., 2005, Calvert, 2000, Demonet et al. 1992, Howard et al. 1992, Price et al. 1996) and the left posterior STS also has been associated

in auditory processing of visually presented letters (Sergent et al. 1992). Learning the meanings of the letters has also been located in STS and MTG areas (Vandenbergeh et al. 1996).

Superior temporal sulcus seems to be connected to several human functions: theory of mind and social perception (Saxe, 2006, Zilbovicius et al., 2006), speech perception (Price, 2000) and also several studies show that the STS is critical for audiovisual integration (e.g. Beauchamp et al., 2004, Cavert, 2001). How does this structure seem to do everything? As a large structure it has been suggested that different parts of the STS are specialized to different functions (Hein & Knight, 2008). In their example, Hein and Knight (2008) describes different regions in the STS: the left STS area hosts speech processing (Wernicke, 1874) and the anterior STS, STG and angular gyrus are involving sentence processing (Dronkers et al., 2004). The MTG has more critical role in speech comprehension. (Dronkers et al., 2004). To take a closer look to this area, it locates on the side of the brain, on the temporal lobe, underneath the Sylvian fissure. *KUVA MISTÄ?*

In this thesis the interest is in audiovisual integration and learning in letter - speech sounds. Like the previous literature, the aim is to try to locate the brain regions and time course where this multisensory integration happens via brain research device, magnetoencephalogram. Because the interest is in time course of brain function, MEG is more natural option for brain research device than for example fMRI, which is commonly used in studies presented in this section.

Magnetoencephalogram, MEG

In this thesis audiovisual integration was studied by magnetoencephalogram, MEG. As a noninvasive technique, MEG detects the magnetic fields outside of the skull, created by neurons. Neurons generate ion currents in the brain and the neuronal communication happens over electric impulses, so-called action potentials. The electric pulse travels along the axon ending up to the axon terminal, where the neurotransmitter molecules are released to synaptic cleft. Neurotransmitters are flown through the gap and attach the second cell, postsynaptic cell membrane. This attachment changes the potential of the second cell, which is called postsynaptic potential, PSP. Because of this event an electric field and current is caused in

postsynaptic cell. This is the current that creates the magnetic field around it which is then detected with MEG device. (Hämäläinen et al., 1993).

MEG recordings are usually sensitive only to the uppermost layer of the brain, the cortex, because the electric signals on the surface brain area can be detected better than the ones coming from deeper areas. Cerebral cortex is approximately 2-4 mm thick and consists of gray matter (Hämäläinen et al., 1993). This gray matter consists of neurons' dendrites, neuronal cell bodies and capillaries (Kolb & Whishaw, 2015). In the gray matter there are mainly two types of cells: pyramidal cells and stellate cells. Stellate cell dendrites are oriented perpendicular to the skull, which means that they are silent to MEG because the magnetic field does not reach outside of the head. Pyramidal cells' dendrites which are oriented tangential in relation to skull are important to MEG signal. When the dendrite receives stimuli from other cells and eventually the postsynaptic potential creates the electric current, the magnetic field can be detected with MEG because of the tangential orientation: the round magnetic field is now possible to exit the head and the MEG device's superconducting quantum interference device (SQUID) can detect the field. (see appendix 1.1.)

The aims and research questions

The aim of this thesis is to study neuropsychological features of audiovisual learning. Chinese language offers a great opportunity to explore speech sounds combined to visual characters because of its very different form from Finnish language. For native Finnish speaker Chinese is also pretty rare language that is not encountered in everyday life. In this study, adult, native Finnish speakers' performance is compared to that of adult native Chinese speakers. For both of the groups, same experiments were performed, except the Chinese group didn't have the training phase like Finnish group did. The idea is to compare, how differently native language speakers vs. non-native speaker brains process the heard speech sounds. The research questions are:

1. Which processing stages are different in Finnish and Chinese subjects? And are these differences observed only in AV condition or also in unimodal conditions?
2. Which brain areas are involved in processing Chinese characters?

Based on previous literature, the hypothesis is to see activation on the STS region. What is new in this research experiment is the fact that Chinese characters doesn't represent a traditional letter to Finnish people, but more like some small picture or figure. In addition, the comparison is performed between a group that knows the meaning of these sound - character pairs and a group that isn't familiar with these pairs.

Exploring the neural mechanisms of audiovisual learning helps us to understand how the brain functions in everyday-tasks. It is also important to understand the basic sensory learning of the brain in order to understand the dysfunctions of the brain, for instance dyslexia, neurological disorders or problems in speech perception.

METHODS

This study was carried out in University of Jyväskylä, department of Psychology. The study is part of a larger project *AV Learning*. The study is funded by EU -projects ChildBrain and Predictable projects, European Training Network, Marie Sklodowska-Curie Actions, and also University of Jyväskylä, Faculty of Social Sciences. The ethical approval for the project has been applied and received from the Ethical Board of University of Jyväskylä, and the research project obeys the declaration of Helsinki. All the participants have given their written consent.

Subjects

The data of this thesis has been collected at the University of Jyväskylä, Psychology department's MEG lab, mainly during the autumn of 2015, and a few subjects were measured at the spring of 2016. The participants were recruited university students and friends of the researchers.

In this study 22 participants were scanned in the MEG-device, half of them native Finnish speakers and another half native Chinese speakers. Eleven Finnish participants were studied, 4 males and 7 females, age between 21 - 32 years, mean 24,72 and SD 3,28 years. In the background survey participants age, handedness, medication, learning difficulties, neurological disorders, permanent head injuries and claustrophobia were found out. None of the subjects had any neurological conditions or used any medication affecting to central nervous system. Subjects were native Finnish speakers and had never studied any Chinese or were not exposed to it in any other way. Some exclusion criteria concerned about the magnetic features of participant: no braces or any other metallic items were allowed to be worn because of the MEG's sensitivity to metal objects.

Chinese group consisted of eleven participant, 2 males and 9 females, age between 19 - 28, mean 23,45 and SD 3,44 years. In the background survey the same issues were found out as in the Finnish participants survey. None of the Chinese subjects had any medication, learning difficulties, neurological disorders, permanent head injuries nor claustrophobia. In return for the 1-2 hour measurement session all of the participants got a free movie ticket (approximate value 10 euros).

Stimuli and experimental design

In this study the participants were presented audiovisual stimuli which included Chinese speech sounds occasionally combined to Chinese character. Participants were asked to sit still in the MEG device and focus on the fixation point shown in the screen to reduce the eye movements. A four-button response box were given to them and they were instructed to use right hand to press the answer button. The experimental design contained three phases:

1) Pre-test
2) Training
3) Post-test

The pre- and post-test were the same. In these phases, Chinese speech sounds and the Chinese characters were played at the same speed to the participant. In the screen (Elekta Neuromag) the fixation point was shown for 1000 ms, then the character and the sound were played either alone or starting at the same time, character lasting for 1000 ms and the sound for approximately 400 ms (403 - 495 depending on the syllable). There were six different characters and six sounds which each occurred 54 times. The used Chinese syllables were /du/ /tu/ /gu/ /ku/ /pu/ /bu/. Occasionally a question mark appeared to the screen asking which of the four options was the character he/she heard or saw before the last one. Question marks appeared randomly by 7,5% chance after every shown stimuli. The task continued after the subject pressed a response button. The audiovisual stimuli varied between the conditions a) visual stimuli alone (V), b) auditory stimuli alone (A) and c) visual and auditory stimuli together, either in congruent way which means the sound and character belonged together, or incongruent way, where the sound and character didn't belong together. Each A and V combination occurred equally many times, meaning that the participants could not use a statistical learning strategy to learn the right sounds and characters.

The subject's task was to remember what was the second-to-last shown character/syllable and in that way keep the participants attention high. After pressing the response button subjects had a feedback on the screen, showing was the answer correct or not and how many percents of correct answers he/she had had of all. The purpose of the feedback was to motivate the participants to try to remember the asked character/syllable.The experimental design illustrated in Table 1.

Table 1. The stimuli shown in participant's screen. After pressing the control button a feedback screen showed whether the answer was correct or not and what was the total score in percentage.

In the training phase, participants were shown Chinese character and sound/syllable pairs. Half of which were always paired in the same way while half were paired randomly. Each pair was shown also 54 times, fixation point lasting for 1000 ms and the character 1000 ms and sound approximately 400 ms. Also in this phase subjects had the feedback in the screen after answering.

The last one, post-test was same as the pre-test. To the Chinese group only the pre/post -phase was performed, Finnish group had all three phases. In this thesis only the pre-test phase was compared between Finnish and Chinese groups. The effect of the training phase and post-test have been studied in other thesis.

Each of the phases lasted approximately 20 minutes, depending how quickly the participant responded to each question mark. All the participants were instructed to answer as quickly as possible. One other experiment was presented to the participants to be reported elsewhere.

Equipment

In this study magnetoencephalogram (MEG) was used as a brain imaging device. For the noise reduction, the MEG device is placed in a sheltered chamber of double walls of aluminium and mu-metal so that the magnetic fields of regular environment will not disturb the weak signals coming from the brain. This MEG device (Elekta Neuromag TRIUX) has whole-head 306 sensor helmet, 102 arrays containing one magnetometer and two orthogonal, planar gradiometers. Five Head Position Indicator (HPI) coils were attached to the scalp of a participant, three on the forehead and two behind the ears as high as possible. After attachment the coils were digitized with Polhemus Fast Track. In addition to the coils, also the specific landmarks were digitized: nasion and left and right preauricular points. Also the head shape was drawn with digitizing pen for the forthcoming analysis phase. The aim of the HPI coils is to monitor the location of the head in relation to sensors during the MEG recording by sending a 293-321 Hz sinusoidal current into the coils. A sampling rate of 1000 Hz and a band-pass filter of 0.3-300 Hz was used in data collection. Two electrodes were used to record eye movement, placed on top corner of the left eye and on the lower corner of the right eye. Ground electrode was attached to the collarbone. The used earpieces were Care Fusion, Model TIP-300. Participant used their right hand to give their answer with the control box.

For the stimulus design, a software Presentation was used to present the stimuli to the participants.

Data processing

The data was preprocessed using MaxFilter software (Elekta Neuromag). The purpose of MaxFilter is to clean the data from the external noise caused by the magnetic fields coming from outside the participant. MaxFilter program was also used to correct the head movements from the data and corrected the head position as it was at the beginning of the measurement session. Temporal extension signal subspace separation, tSSS (Taulu et al., 2005) in the MaxFilter program was performed to clean the data.

After the MaxFilter treatment the data was processed using BESA Research 6.1 software (BESA GmbH, Gräfelfing, Germany), where the data was averaged and then performed a distributed source localization. At first, the blinks were removed with ICA tool (infomax algorithm, xx sec long MEG segment), which recreates the data without the blinks. This was performed to both gradiometers and magnetometers. After the blink corrections the artefact scanning and averaging was performed. The artefact scanning showed the epochs which amplitude was over 4000 fT and those were rejected from the averaging. The averaged time window was -100 ms pre-stimulus and 800 ms post-stimulus. The baseline was set as -100 - 0 ms before stimulus onset. For the Finnish group, there were 102 (94,94%), 103 (95,87%), 100 (93,43%), 102 (94,61%), 50 (94,27%), 50 (92,59%), 51 (94,27%) and 51 (94,94%) accepted epochs averaged in conditions auditory only, visual only, AV congruent, AV incongruent, AV learnable congruent, AV unlearnable congruent, AV learnable incongruent and AV unlearnable incongruent. For the Chinese, corresponding accepted epochs were 93 (86,44%), 85 (78,36%), 93 (86,27%), 29 (85,43%), 46 (85,85%), 47 (87,71%), 46 (85,85%) and 46 (85,01%). In this thesis, the four first conditions, V, A, AV congruent and AV incongruent had a further analysis.

When the data cleaning and averaging was done, the source localization was performed to the data. For the examination of the signal, Classical LORETA Recursively Applied (CLARA; Hoechstetter et al., 2010) method was used. The low pass filter was set as 40 Hz. CLARA - source modelling analysis was performed on all of the subjects, and for both of the groups analysis time windows were set for 50-200 ms and 200-500 ms. These time windows were set because of the interest was in the sensory responses in the brain (50-200 ms) but also in the

longer term reaction (200-500 ms). Spherical head model was selected as a basis for the source analysis.

Statistics

As a statistic software, BESA Statistics 2.0 was used to perform statistical analysis. Because two different groups were compared, Finnish and Chinese, the unpaired two tailed t-test was selected. When the comparison happened between conditions (i.e., congruent vs. incongruent audio-visual stimuli) the paired t-test was selected.

In unpaired t-test, Finnish and Chinese groups were compared in visual only, auditory only, audiovisual congruent and audiovisual incongruent situations, in time windows 1 and 2, making 8 studied situations. In paired t-test four possible situations were studied: Finnish group in audiovisual congruent and incongruent conditions in time windows 1 and the same conditions in time window 2. Also Chinese group was compared in situations audiovisual congruent vs. incongruent, in both 1 and 2 time windows.

After the statistical analysis, the results were significant in several conditions. However, a large cluster appeared on the top of the head in many cases, which did not represent an actual brain activation. To avoid the cluster and in order to examine the results more reliable, a specific brain coordinates were used in the second statistical analysis. On the grounds of previously literature, visual cortex, left and right pSTS, left and right auditory area / Heschl's gyrus and left and right fusiform areas were selected to this analysis. The coordinates were as follows:

Region	x	y	z
Visual cortex	7	-68	8
pSTS left	-53	-31	0
pSTS right	48	-31	6
Heschl's gyrus left	-48	-20	8
Heschl's gyrus right	48	-20	8
Fusiform left	-42	-57	-6
Fusiform right	42	-57	-6

Table 1. The used coordinates in the second statistical analysis.

However, the statistic analysis on these coordinates were performed on those conditions, which showed significance in the first statistical analysis. The conditions were visual only in time window 1, auditory only in time window 2, congruent in time window 1 and incongruent in time window 1. In addition to these conditions, all of the four paired -conditions showed significance and therefore the special coordinates were also used to them. Accordingly, for the eight conditions were re-computed statistical analyses in the seven coordinates from which were interested in.

RESULTS

The aim of this thesis was to find out, which brain regions attend processing the speech sounds and characters and does it affect on the processing, when the condition is unimodal or bimodal. In addition, the interest was also in the difference how did the processing look in the brains which were familiar with the stimulus pairs (Chinese) or which weren't (Finnish).

Here are presented the results of the critical coordinates which were used to avoid the large cluster appearing in several conditions.

In the first significant condition, visual only in time window 1, visual coordinate (Belliveau et al., 1991) showed under 0.05 significance.

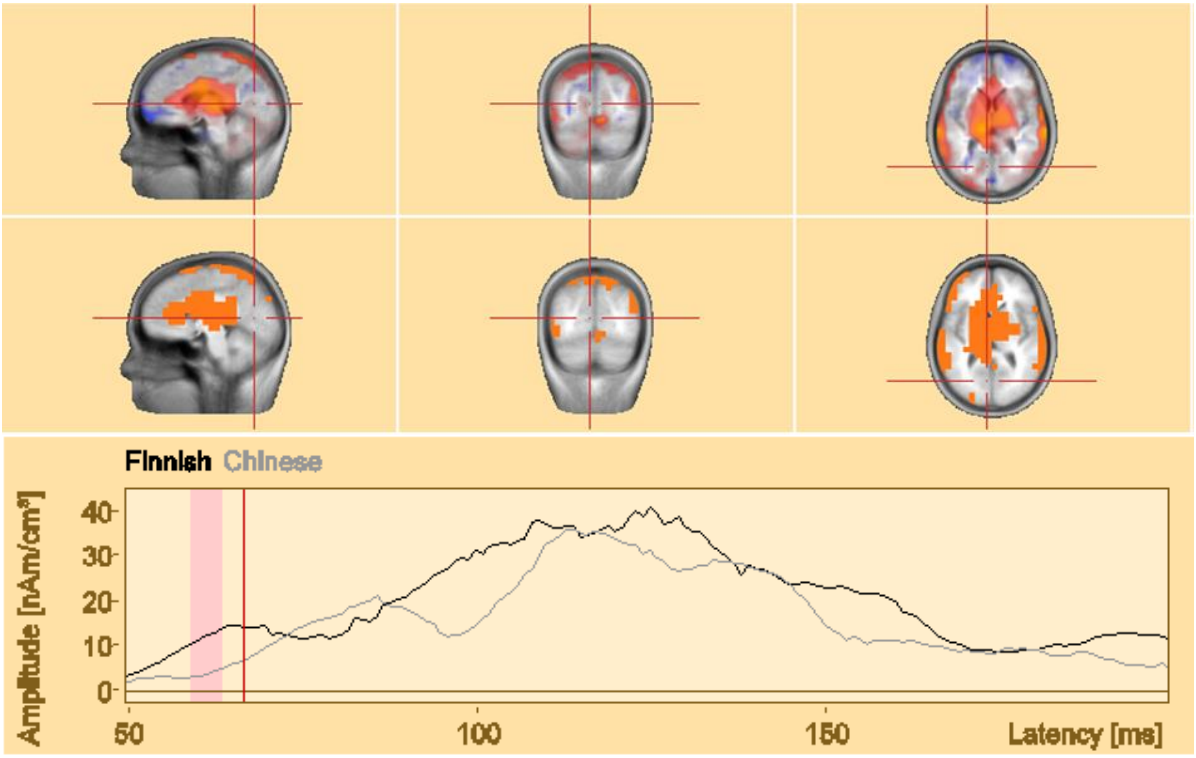

Figure 1. Top: t-values for the group comparison. Middle: statistically significant clusters. Bottom: source amplitude values at the crosshair position. When the participant was shown only visual stimuli, chinese characters, Finnish groups brain activation in the visual cortex was shown stronger activation than Chinese. The response developed in the first time window, during 59-64 ms.

The third condition, congruent time window 1, showed under 0.05 significance in right and left fusiforms coordinates (Talairach)

Right:

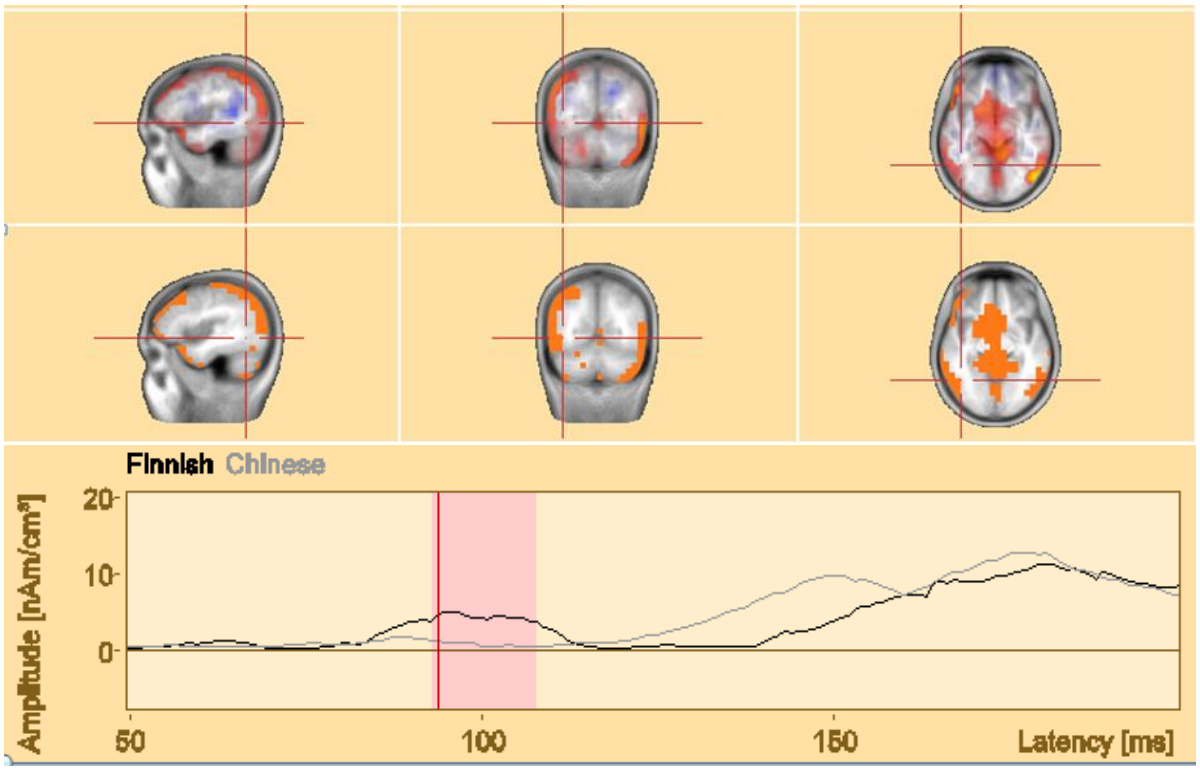

Figure 2. When stimuli were presented congruently, Finnish groups showed greater activation in the right fusiform area during 93 - 108 ms.

Left:

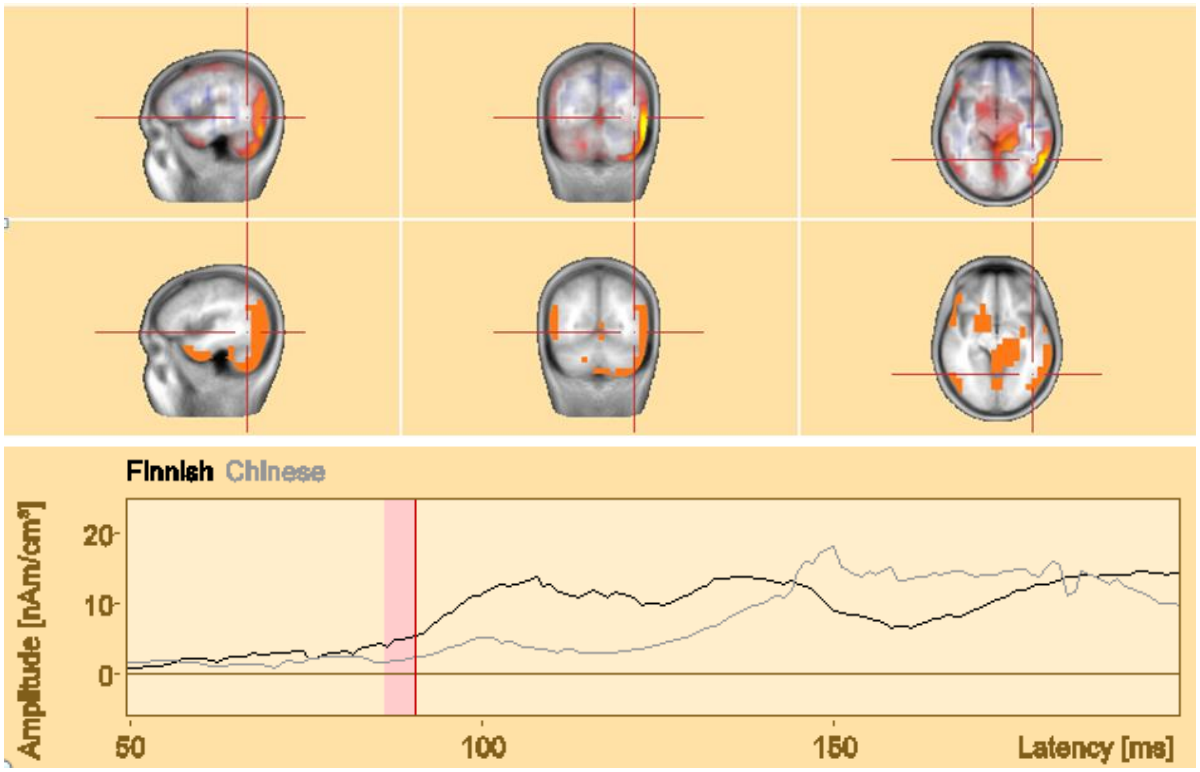

Figure 3. On the left fusiform, Finnish groups showed stronger activation in the first time window, during 86 - 91 ms.

The second and the fourth condition, auditory only in time window 2 and incongruent in time window 1, didn't show any significances in any given coordinates (Raij et al., 2000, Talairach).

The seven coordinates were also performed to the paired tests. In all four conditions, none of the coordinates showed any significance.

DISCUSSION

Activation in the cortex: Discovered differences between the groups and conditions

The aim of this thesis was to study where audiovisual integration in the brain is processed and whether knowing the meaning of the speech sound - symbol pairs had any effect on the processing. The participants were played Chinese speech sounds and characters either within pairs or alone and the subject's task was to remember what was the second to last shown character or sound. Two groups of participants were formed: native Chinese speakers and Finnish participants, who were not familiar with Chinese. Participants' brain activation was recorded by magnetoencephalogram, MEG.

In the analysis, source localization was performed by BESA -software, CLARA -algorithm. BESA Statistics was used to compare the two groups and conditions between the groups. Statistic calculation was also performed to specific coordinates which represented the regions of interest: visual cortex, right and left pSTS, right and left auditory cortex / Heschl's gyrus and left and right fusiform gyrus. The research questions considered which processing stages are different in Finnish and Chinese participants, and are these differences seen only in AV condition or also in unimodal conditions. In addition, the brain areas involving in the processing of Chinese characters was an area of interest.

According to the previous literature, it would have been conventional to find some activation in the pSTS areas. However, the results didn't show any activation on these areas but in the fusiform gyrus. In unpaired tests, the results showed that Finnish group's visual cortex activated stronger than Chinese group's in visual only condition, time window 1. Also, Finnish group's both right and left fusiform areas were more powerfully activated in congruent condition, time window 1, than the Chinese. In paired tests no differences between the conditions were found. When it comes to research questions, the processing between Finns and Chinese do differ and especially in visual level, and it can be observed both in unimodal and bimodal conditions. The activated brain regions were areas known to relate visual processing, visual cortex and fusiform gyrus.

The results are especially interesting because they are rather conflicting to the previous researches. Like many other studies, audiovisual processing has been located several times in pSTS areas (Beauchamp et al., 2004, Gilbert et al., 2013, Tanabe et al., 2005). Contrary to expectations, neither of the Finnish or Chinese groups had significant activation on these areas.

The role of fusiform gyrus in the speech sound - letter processing

This is not the first time when the fusiform gyrus is reported to take an action in word recognition or processing. For example, Binder, Medler, Westbury, Liebenthal and Buchanan (2006) reported in their functional magnetic resonance imaging (fMRI) study that the left lateral fusiform gyrus responded to both familiar and unfamiliar sequences of letters. They suggest that the fusiform gyrus is responsible of perceptual tuning, when it comes to orthographic processing. Still, they did not have any auditory stimuli in their experiment, so it doesn't directly connect with these results here, when the fusiforms were activated on AV conditions.

Fusiform has been demonstrated as an area which first adopts a role in mapping print and sound, and is later crucial part of the reading network (Brem, Bach, Kucian, Kujala, Guttorm, Martin, Lyytinen et al., 2010). In their report, Brem et al. (2010) found activation in the left occipito-temporal cortex, here referred as visual-word-forming-system, VWFS (Cohen et al., 2000), which is a part of the fusiform gyrus. After a brief training period, nonreading kindergarden children were scanned in fMRI, which showed activation in fusiform gyrus (including other regions) when words were presented in V, A and VA trials.

When it comes to temporal qualities of processing in the fusiform gyrus the results here show rather early development of the response. Both right and left hemispheres were activated mostly under 100 ms, right fusiform in 93 - 108 ms and left 86 - 91 ms. Most of the fusiform related researches has been performed by fMRI or positron emission tomography (PET) where the temporal resolution isn't that reliable (e.g. Binder at al., 2006). However, Cohen et al. (2000) reported their ERP's to peak in 180 - 200 ms post-stimulus onset. This also makes the results of this thesis quite conflicting.

Some studies strongly suggest that VWF system in the mid-portion on the left fusiform gyrus activates whenever a literate person is reading words (Cohen et al. 2000, Cohen et al. 2002), and some studies do not consider this area specially a word forming region, but it participates in several functions (Price & Devlin, 2003). However, these results in this thesis the fusiform gyrus, both of them, participated when processing non-familiar Chinese character, which didn't include letter strings. Also, these characters were combined to specific speech sound. Accordingly, word forming area or not, this structure involves when processing unknown visual

and auditory stimulus together. These results probably raise more questions about the role of the fusiform gyrus in the processing of audiovisual stimuli than actually gave answers.

As a conclusion, it appears that the activation of the fusiform gyrus is linked to forming an early networks which are in charge of reading, among other reading-like functions. It could be possible, that Finnish subject's brains started to process Chinese characters as any other letter strings and started to form a link between the seen character and the heard speech sound, and the processing happened in the fusiform gyrus. Also, the results indicate that the Finnish group reacted to this new kind of V and AV condition especially by processing in on the visual-related brain regions. Could this be explained by the unusual form of the Chinese characters which aren't usually seen in Finnish environment? On the other hand, the auditory cortex did not show any significant response to Chinese speech sounds. Could it be related to the fact that the used Chinese speech sounds didn't differ much of a regular Finnish pronunciation, so the extra processing on the auditory cortex wasn't needed? To recognize the unfamiliar character is much challenging than recognize the speech sounds, in which case visual processing is required more than auditory processing.

Critique

In this thesis, also unusual results were reported. The exceptional results here are the difference between the AV congruent and AV incongruent conditions. Because this thesis considered only the pre-test part for Finnish group, the subjects did not have any training to Chinese character nor speech sounds. Accordingly, it means that the Finnish subjects would not know whether the AV pairs would be presented correctly or not. The strange part is, that only the AV congruent condition showed significance between the groups, not incongruent condition. Also notable thing is, that the Chinese groups processing did not differ, whether the condition was congruent or not, although the Chinese group knew the meaning of the characters and speech sounds: it did not show in processing level whether the pairs were correct or not.

As a critique can be said, that not all participants were analysed in this theses. After analysing these participants presented here, many new subjects were scanned but they weren't added to this thesis. This could also be explaining some of the unusual results described above. A greater *n* could have affected results making them more reliable and further refining the statistically significant areas of the brain.

The results can also be criticized from the narrowness of the times showing significance. It has to be taken into consideration that the smallest, only few milliseconds lasting time windows can also originate from coincidence.

Further study

Audiovisual processing has been studied widely but the conflicting results reported here show that the location of the audiovisual processing isn't still clear. Also many different features can be studied of audiovisual integration: the AV stimulus pairs consisting of speech sound - letter pairs, or the letter strings forming a real word or a pseudo word, or the AV pairs can be something completely different, for example objects and sounds. It is yet to be found, whether all these AV pairs are processed in same location or does there exist some own regions, like the VWF systems. Also the temporal qualities of the audiovisual processing are studied. Most of the studies presented this thesis are fMRI studies (e.g. Beauchamp et al., 2004, Tanabe et al., 2005) which don't have very good temporal resolution. For this reason, new and multifaceted research is needed to solve the mystery of audiovisual processing in the brain, which is behind all human function.

Acknowledgements

Greatest compliments to the supervisor Jarmo Hämäläinen who had always time for my questions and concerns, especially in the data processing phase.

REFERENCES:

Amunts, K., Lenzen, M., Friederici, A. D., Schleicher, A., Morosan, P., Palomero-Gallagher, N., & Zilles, K. (2010). Broca's region: novel organizational principles and multiple receptor mapping. *PLoS Biol*, *8*(9), e1000489.

Beauchamp, M. S., Lee, K. E., Argall, B. D., & Martin, A. (2004). Integration of auditory and visual information about objects in superior temporal sulcus.*Neuron*, *41*(5), 809-823.

Belliveau, J. W., Kennedy, D. N., McKinstry, R. C., Buchbinder, B. R., Weisskoff, R. M., Cohen, M. S., ... & Rosen, B. R. (1991). Functional mapping of the human visual cortex by magnetic resonance imaging. *Science*,*254*(5032), 716-719.

Binder, J. R., Medler, D. A., Westbury, C. F., Liebenthal, E., & Buchanan, L. (2006). Tuning of the human left fusiform gyrus to sublexical orthographic structure. *Neuroimage*, *33*(2), 739-748.

Bolognini, N., Frassinetti, F., Serino, A., & Làdavas, E. (2005). "Acoustical vision" of below threshold stimuli: interaction among spatially converging audiovisual inputs. *Experimental brain research*, *160*(3), 273-282.

Brem, S., Bach, S., Kucian, K., Kujala, J. V., Guttorm, T. K., Martin, E., ... & Richardson, U. (2010). Brain sensitivity to print emerges when children learn letter–speech sound correspondences. *Proceedings of the National Academy of Sciences*, *107*(17), 7939-7944.

Calvert, G. A. (2001). Crossmodal processing in the human brain: insights from functional neuroimaging studies. *Cerebral cortex*, *11*(12), 1110-1123.

Calvert, G. A., & Thesen, T. (2004). Multisensory integration: methodological approaches and emerging principles in the human brain. *Journal of Physiology-Paris*, *98*(1), 191-205.

Calvert, G. A., Campbell, R., & Brammer, M. J. (2000). Evidence from functional magnetic resonance imaging of crossmodal binding in the human heteromodal cortex. *Current biology*, *10*(11), 649-657.

Cohen, L., Dehaene, S., Naccache, L., Lehéricy, S., Dehaene-Lambertz, G., Hénaff, M. A., & Michel, F. (2000). The visual word form area. *Brain*, *123*(2), 291-307.

Cohen, L., Lehéricy, S., Chochon, F., Lemer, C., Rivaud, S., & Dehaene, S. (2002). Language-specific tuning of visual cortex? Functional properties of the Visual Word Form Area. *Brain*, *125*(5), 1054-1069.

Demonet, J. F., Chollet, F., Ramsay, S., Cardebat, D., Nespoulous, J. L., Wise, R., ... & Frackowiak, R. (1992). The anatomy of phonological and semantic processing in normal subjects. *Brain*, *115*(6), 1753-1768.

Dronkers, N. F. (1996). A new brain region for coordinating speech articulation.*Nature*, *384*(6605), 159-161.

Dronkers, N. F., Wilkins, D. P., Van Valin, R. D., Redfern, B. B., & Jaeger, J. J. (2004). Lesion analysis of the brain areas involved in language comprehension.*Cognition*, *92*(1), 145-177.

Frassinetti, F., Bolognini, N., & Làdavas, E. (2002). Enhancement of visual perception by crossmodal visuo-auditory interaction. *Experimental Brain Research*, *147*(3), 332-343.

Froyen, D., Willems, G., & Blomert, L. (2011). Evidence for a specific cross-modal association deficit in dyslexia: an electrophysiological study of letter–speech sound processing. *Developmental science*, *14*(4), 635-648.

Gilbert, J. R., Pillai, A. S., & Horwitz, B. (2013). Assessing crossmodal matching of abstract auditory and visual stimuli in posterior superior temporal sulcus with MEG. *Brain and cognition*, *82*(2), 161-170.

Gómez, R. L., & Gerken, L. (2000). Infant artificial language learning and language acquisition. *Trends in cognitive sciences*, *4*(5), 178-186.

Hämäläinen, M., Hari, R., Ilmoniemi, R. J., Knuutila, J., & Lounasmaa, O. V. (1993). Magnetoencephalography—theory, instrumentation, and applications to noninvasive studies of the working human brain. *Reviews of modern Physics*,*65*(2), 413.

Hashimoto, R., & Sakai, K. L. (2004). Learning letters in adulthood: direct visualization of cortical plasticity for forming a new link between orthography and phonology. *Neuron*, *42*(2), 311-322.

Hein, G., & Knight, R. T. (2008). Superior temporal sulcus—it's my area: or is it?. *Journal of cognitive neuroscience*, *20*(12), 2125-2136.

Herath, P., Klingberg, T., Young, J., Amunts, K., & Roland, P. (2001). Neural correlates of dual task interference can be dissociated from those of divided attention: an fMRI study. *Cerebral cortex*, *11*(9), 796-805.

Hertrich, I., Mathiak, K., Lutzenberger, W., Menning, H., & Ackermann, H. (2007). Sequential audiovisual interactions during speech perception: a whole-head MEG study. *Neuropsychologia*, *45*(6), 1342-1354.

Hoechstetter, K, Berg, P., Scherg, M. (2010). BESA Research Tutorial 4: Distributed Source Imaging.

Howard, D., Patterson, K., Wise, R., Brown, W. D., Friston, K., Weiller, C., & FRACKOWIAK, R. (1992). The cortical localization of the lexicons. *Brain*,*115*(6), 1769-1782.

Iacoboni, M., & Wilson, S. M. (2006). Beyond a single area: motor control and language within a neural architecture encompassing Broca's area. *Cortex,42*(4), 503-506.

Institute for Learning and Brain Science, Washington University. (2012) http://ilabs.uw.edu/what-magnetoencephalography-meg

Kolb, B., & Whishaw, I. Q. (2015). *Fundamentals of human neuropsychology*. Macmillan.

Mechelli, A., Gorno-Tempini, M. L., & Price, C. J. (2003). Neuroimaging studies of word and pseudoword reading: consistencies, inconsistencies, and limitations. *Journal of cognitive neuroscience*, *15*(2), 260-271.

Novick, J. M., Trueswell, J. C., & Thompson-Schill, S. L. (2005). Cognitive control and parsing: Reexamining the role of Broca's area in sentence comprehension. *Cognitive, Affective, & Behavioral Neuroscience*, *5*(3), 263-281.

Paulesu, E., McCrory, E., Fazio, F., Menoncello, L., Brunswick, N., Cappa, S. F., ... & Pesenti, S. (2000). A cultural effect on brain function. *Nature neuroscience*, *3*(1), 91-96.

Price, C. J. (2000). The anatomy of language: contributions from functional neuroimaging. *Journal of anatomy*, *197*(03), 335-359.

Price, C. J., & Devlin, J. T. (2003). The myth of the visual word form area.*Neuroimage*, *19*(3), 473-481.

Price, C. J., Wise, R. J., & Frackowiak, R. S. (1996). Demonstrating the implicit processing of visually presented words and pseudowords. *Cerebral cortex*, *6*(1), 62-70.

Raij, T., Uutela, K., & Hari, R. (2000). Audiovisual integration of letters in the human brain. *Neuron*, *28*(2), 617-625.

Saxe, R. (2006). Uniquely human social cognition. *Current opinion in neurobiology*, *16*(2), 235-239.

Tanabe, H. C., Honda, M., & Sadato, N. (2005). Functionally segregated neural substrates for arbitrary audiovisual paired-association learning. *The Journal of neuroscience*, *25*(27), 6409-6418.

Taulu, S., Simola, J., & Kajola, M. (2005). Applications of the signal space separation method. *IEEE transactions on signal processing*, *53*(9), 3359-3372.

Tuomainen, J., Andersen, T. S., Tiippana, K., & Sams, M. (2005). Audio–visual speech perception is special. *Cognition*, *96*(1), B13-B22.

Van Atteveldt, N., Formisano, E., Goebel, R., & Blomert, L. (2004). Integration of letters and speech sounds in the human brain. *Neuron*, *43*(2), 271-282.

Vandenberghe, R., Price, C., Wise, R., Josephs, O., & Frackowiak, R. S. (1996). Functional anatomy of a common semantic system for words and pictures. *Nature*, *383*(6597), 254-6.

Willems, R. M., Özyürek, A., & Hagoort, P. (2007). When language meets action: The neural integration of gesture and speech. *Cerebral Cortex*, *17*(10), 2322-2333.

Zilbovicius, M., Meresse, I., Chabane, N., Brunelle, F., Samson, Y., & Boddaert, N. (2006). Autism, the superior temporal sulcus and social perception. *Trends in neurosciences*, *29*(7), 359-366.