

UNIVERSITY OF OSNABRUECK

BACHELOR THESIS

The Functional Neuroanatomy of Speech Perception

Insights from Electrocorticography on an
Epilepsy Patient Watching a Movie

PHILIPP KUHNKE

1st Supervisor:

PROF. DR. JUTTA L. MUELLER
(University of Osnabrueck)

2nd Supervisor:

PROF. DR. PETER BOSCH
(University of Osnabrueck)

Research Supervisor:

PROF. DR. GABRIEL KREIMAN
(Harvard University)

July 15, 2014

There are several people I want to thank for supporting me in the compilation of this ambitious thesis.

First of all, of course, my first supervisor Prof. Dr. Jutta L. Mueller who went out of her way to answer any question I had and to get rid of any problem I was faced with, dealing with the uncontrolled nature of the stimulus. I will not take this for granted, especially considering the other things (or rather tiny humans) she had to deal with during this time. Her profound knowledge of neurolinguistic research has helped me tremendously.

Secondly, I have to thank my professors of linguistics, my second supervisor Prof. Dr. Peter Bosch and Harvard's Prof. Dr. Maria Polinsky. Ever since the first sessions of the "Introduction to Computational Linguistics", Peter Bosch has been one of the main reasons why I am fascinated by the study of language. He manages to deal with the most challenging issues in theoretical, computational, as well as psycho- and neurolinguistics with inspiring sophistication, wit and most importantly passion. It has been an incredible pleasure to be a tutor for the very same introductory course in linguistics the following two years. I was a student of Maria Polinsky during my time at Harvard where she did the course "Language, Structure and Cognition". While I knew a lot of the material already from Peter Bosch's lectures, the different perspective of a syntactician incorporating methods from cognitive science into her research was highly refreshing. What was most inspiring, however, were her lab meetings in Boylston Hall. Right next to the office of Gennaro Chierchia, I listened to lively discussions about current issues in theoretical linguistics while eating Pizza. This wonderful atmosphere in the company of wonderful people inspired me once again that I want to pursue a career in this field.

Then, I owe a debt of gratitude to everyone in Prof. Dr. Kreiman's lab: Hanlin Tang and Dr. Radhika Madhevan who became dear friends to me and especially Dr. Jed Singer who was my main "partner" for the MATLAB work I did in Boston. The help he provided on the technical aspects was incredible and I could not thank him enough. I also want to say "thanks" to Evelina Fedorenko (whom I was lucky enough to just call "Ev"): During the first two to three months in Boston of struggling with the difficult nature of the stimulus, trying to figure out what kinds of questions can be validly asked and actually answered using the data, she was the one that provided me with the ideas which pushed me into the right direction.

Finally, I of course want to thank Prof. Dr. Gabriel Kreiman himself. From the first moment we met, he treated me as someone who was capable of doing actual scientific research on his own. Instead of giving me some menial task of some other person's project, he encouraged me to start my own individual research project. I remember Gabriel's words very well: "So we have this data from a patient who watched a movie, but we've never worked on language. Are you interested?". Looking back, I have to thank Dr. Kreiman for letting me figure everything out myself. Although it was difficult, going through this experience has taught me so much about research and perseverance.

Last but not least, I have to thank my parents. I owe it to them that I feel at home in the academic culture. They have strongly supported me in my academic endeavors as well as in every other aspect of life.

ABSTRACT

The processing of heard speech in the brain comprises several stages: After an initial analysis of all sounds, speech and non-speech, in the primary auditory cortex (PAC), speech is identified and processed differently from non-speech sounds to determine its linguistic content. Previous findings show that language-specific processes involve a left-lateralized network of cortical regions in the temporal and inferior frontal lobes.

Speech perception in natural face-to-face communication also involves seeing and "reading" the lips of the speaker. This can strongly influence speech comprehension, which indicates that auditory and visual speech information are integrated. It has been shown that lip reading essentially engages the same neural network as auditory speech.

In this thesis, an electrocorticography (ECoG) experiment is presented which uses a movie as the stimulus to come close to real life conditions of speech perception. The electrophysiological data is analyzed with respect to two questions: Which brain regions distinguish (1) speech vs. non-speech sounds; and which regions differentiate between (2) speech with seen lip movements of the speaker vs. speech without seen lip movements. The results demonstrate that regions sensitive to contrast (1) are mostly regions involved in speech processing stages beyond the PAC; exceptions are the lingual gyrus, superior parietal lobe and early visual cortex. Areas sensitive to the contrast in (2) include the left STS, the presumptive location of audiovisual integration. Most of the other regions differentially activated for this contrast are speech processing regions. This illustrates that visual speech significantly alters processing in the "auditory" speech processing regions. The time course of differential activations reveals striking discrepancies in differentiation times of different electrodes in the same region (e.g. MTG or precuneus), which suggests that different subregions of the respective areas are involved in distinct functional networks. Thus, a functional subdivision of these regions seems necessary.

CONTENTS

1. <i>Introduction</i>	6
2. <i>Speech Processing in the Brain</i>	7
2.1 Evidence from Aphasias	8
2.1.1 Broca's Aphasia	8
2.1.2 Wernicke's Aphasia and TSA	8
2.2 The Wernicke-Geschwind Model	9
2.3 Problems of the Wernicke-Geschwind Model	10
2.4 Stages in Speech Perception	12
2.4.1 Initial Acoustic-Phonetic Analysis	13
2.4.2 Dorsal and Ventral Streams in Speech Processing	15
2.4.3 Auditory Sentence Processing	16
2.4.4 A Model of the Functional Neuroanatomy of Auditory Sentence Processing	19
2.4.5 Discourse Processing	20
2.5 Audiovisual Speech Perception	21
2.5.1 Behavioral Studies	21
2.5.2 Neuroimaging Studies	22
3. <i>Experiment</i>	25
3.1 Introduction	25
3.2 Methods	25
3.2.1 Participant	25
3.2.2 Apparatus	25
3.2.3 Stimulus Material	29
3.3 Analysis	29
3.4 Results	31
3.4.1 Speech vs. Non-Speech Sounds	31
3.4.2 Speech with Seen Lips vs. Speech without Seen Lips	34
3.5 Discussion	37
3.5.1 Speech vs. Non-Speech Sounds	37
3.5.2 Speech with Seen Lips vs. Speech without Seen Lips	39
3.5.3 Time Course	40
4. <i>Conclusion</i>	42
5. <i>Appendix</i>	44
5.1 Electrode Plots	44
5.1.1 Speech vs. Non-Speech Sounds	45
5.1.2 Speech with Seen Lips vs. Speech without Seen Lips	53

5.2	Abbreviations	61
5.2.1	Brain Regions	61
5.2.2	Imaging Techniques	61
5.2.3	Miscellaneous	62
5.3	Algorithms	62
5.3.1	Statistical Test	62
5.3.2	Loudness and Spectral Frequency Controls	63
5.3.3	Threshold for Period of Significance	65
5.3.4	Comparison of Electrophysiological Responses	67

1. INTRODUCTION

One of the most important and at the same time astonishing capabilities of our cognitive system is to make sense of heard speech as quickly as it does, in a process known as "speech perception". This is especially remarkable since we hear sentences every single day that we have never heard before.

To achieve this, the auditory system of the brain has to single out speech sounds from the auditory input stream, i.e. it has to distinguish speech from non-speech sounds. Then, language-related processes have to analyze the linguistic structure encoded by these sounds in order to enable the comprehension of the meaning of words, novel sentences and even entire stories.

In face-to-face communication, an essential component to speech perception is seeing and thereby "reading" the speaker's lips. We will see that this can significantly influence and enhance speech comprehension. Speech perception, therefore, often also involves the multisensory integration of auditory and visual information.

How does the brain approach these tasks to make speech perception as effortless and automatic as we perceive it to be? What are the neural mechanisms and networks underlying speech perception?

In this thesis, I will approach these issues in an experiment using the stimulus of a movie. While harder to control for confounding factors than traditional lab-prepared stimuli, a movie approximates natural, real life conditions of perceiving speech. The electrophysiological data obtained in this experiment will be used to tackle the two questions introduced above: Which regions in the brain distinguish (i.e. show a significant difference in their responses) (1) speech from non-speech sounds; and which regions distinguish (2) speech with seen lip movements of the speaker from speech without seen lips.

I will begin with a comprehensive review of the literature on speech processing in the brain in chapter 2. In chapter 3, I shall present the experiment conducted for this thesis and discuss its results. I will end with a brief conclusion of the main findings in chapter 4.

2. SPEECH PROCESSING IN THE BRAIN

How does the brain accomplish the task of speech perception? When and how does it process speech sounds differently than non-speech sounds? In this section, I will summarize the literature on speech processing in the brain.¹ Fig. 2.1 shows the most important brain regions in this context.

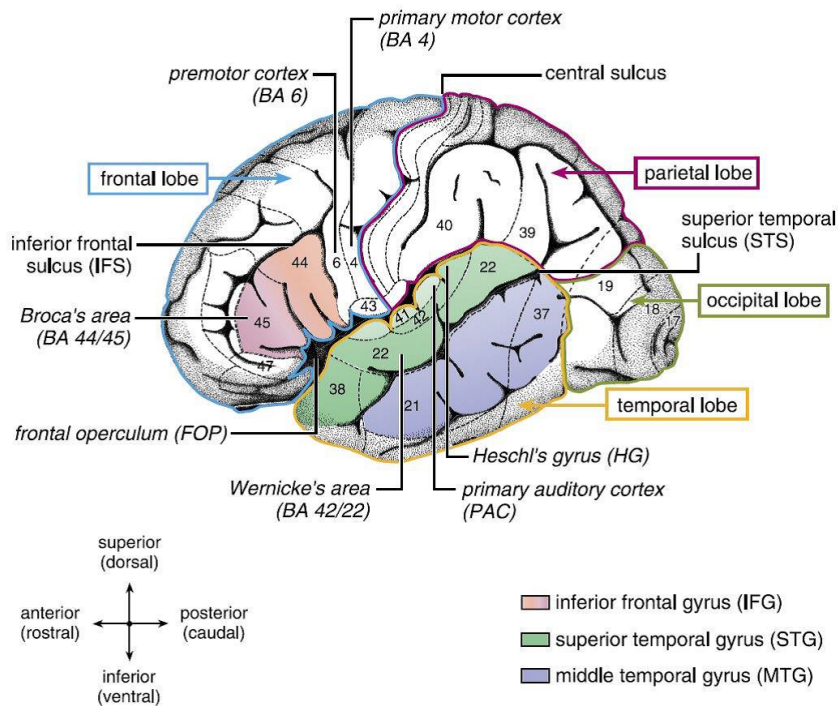


Fig. 2.1: Language-relevant areas of the cerebral cortex (adapted from Friederici 2011)

¹ In order to establish a complete understanding of the neural system underlying speech perception, I shall also include evidence on the involvement of brain areas that are not covered by the patient in my experiment. Specifically, the electrode grid implanted in the patient's brain covered portions of the temporal, parietal and occipital lobes of the brain, but not the frontal lobe. I am convinced, however, that it is still worthwhile to present the role of the frontal lobe in speech perception. Nevertheless, note that for the purpose of my study, the temporal lobe will be the critical cortical structure.

2.1 Evidence from Aphasias

The first hypotheses concerning speech processing in the brain stem from lesion studies, i.e. insights from damage in certain regions of the brain that is associated with specific language deficits called "aphasias".

2.1.1 Broca's Aphasia

Broca's aphasia (also called "expressive aphasia") results from damage to "Broca's area", anatomically defined as comprising Brodmann areas (BA) 44 and 45 in the inferior frontal lobe of the left hemisphere² (Amunts et al., 1999; Broca, 1861). Broca's area is assumed to be responsible for the production of speech for damage to Broca's area causes speech production deficits. Broca's aphasics show great difficulty pronouncing the phones (speech sounds) constituting a word. Consider the following example due to O'Grady et al. (1997, p. 465): a Broca's aphasic trying to say (1a) might instead utter (1b).

- (1) a. It's hard to eat with a spoon.
 b. ... har eat ... wit ... pun.³

While it is tempting to conclude that Broca's area is merely concerned with speech production, many studies have shown that it is involved in language processing as well. I will elaborate on this idea later.

2.1.2 Wernicke's Aphasia and TSA

Wernicke's aphasia and transcortical sensory aphasia (TSA) are directly related to speech comprehension deficits (Damasio, 1992).

Wernicke's aphasia (also called "receptive aphasia") is caused by damage to "Wernicke's area", comprising BA 22 and 42, part of the posterior superior temporal gyrus (STG), in the left hemisphere (Wernicke, 1874). Wernicke's aphasics suffer from severe speech comprehension difficulties. On the other hand, speech production is typically intact. However, the produced utterances lack coherence. As an example, consider the following conversation between an examiner (E) and a Wernicke's aphasic (A) (adopted from (O'Grady et al., 1997, p. 468)):

E: How are you today, Mrs. A?

A: Yes.

E: Have I ever tested you before?

A: No. I mean I haven't.

E: Can you tell me what your name is?

A: No, I don't I... right I'm right now here.

E: What is your address?

A: I cud if I can help these this like you know... to make it. We are seeing for

² Many brain regions related to language processes are lateralized to the left hemisphere for most right-handed people. However, for about 30 percent of left-handed people, language processing regions are often lateralized to the right or even both hemispheres (Stephan and Walter, 2013, p. 66). Throughout this thesis, statements on lateralization will be based solely on findings from right-handed patients in order to avoid confusion. Note, however, that this leaves out a significant portion of the population.

³ The dots (...) represent periods of silence in the utterance.

him. That is my father.

There is an obvious difference in the types of errors between Broca's and Wernicke's aphasics. In contrast to Broca's aphasics, Wernicke's aphasics have no difficulty producing well-formed words, sometimes even well-formed sentences (e.g. "That is my father"). However, most of the time, different utterance segments are completely unrelated to one another semantically. While this example shows how Wernicke's aphasics randomly combine different sentences or even words, there are even more severe cases of Wernicke's aphasia where not only phrases or words but even phonemes are randomly selected (O'Grady et al., 1997, p. 468). Crucially, Wernicke's aphasics are unable to repeat heard speech (Hickok and Poeppel, 2000).

Wernicke's aphasia gives us great insight into language comprehension. It is primarily viewed as a language comprehension deficit. However, as illustrated above, comprehension plays a critical role in production as well. Wernicke's aphasics cannot express their thoughts coherently as they cannot understand their own utterances. Thus, in order to produce coherent language, one has to be able to understand coherent language.

Transcortical Sensory Aphasia

In contrast to Wernicke's aphasics, TSA patients can repeat speech produced by others (Hickok and Poeppel, 2000). Therefore, phonemic perception processes are intact in TSA. The comprehension deficit appears to originate from impairment at a post-phonemic processing stage. TSA is typically associated with damage to areas in the posterior inferior temporal lobe (pITL) including the middle temporal gyrus (MTG) and inferior temporal gyrus (ITG) (Hickok and Poeppel, 2004; Kertesz et al., 1982). The deficits associated with TSA, thus, suggest that areas in the pITL are involved in post-phonemic stages in the processing of speech (Hickok and Poeppel, 2000).

It has been demonstrated that Wernicke's aphasics commit phonemic and semantic errors (Baker et al., 1981). This suggests that Wernicke's aphasia comprises both phonemic and post-phonemic processing impairments (Hickok and Poeppel, 2000). The superior temporal lobe is the area showing the least overlap in lesions associated with Wernicke's aphasia and TSA. For this reason, it is likely to be the cortical substrate of the phonemic processing stage (Damasio, 1998).

However, overall, the impairment in phonemic perception in Wernicke's aphasia does not appear to be severe (Hickok and Poeppel, 2000). For instance, Baker et al. (1981) conducted a word-to-picture matching task with Wernicke's aphasics, who chose the correct phonological representation of the target picture 80% of the time. In conclusion, whereas a phonemic perception deficit seems to contribute to speech comprehension deficits in Wernicke's aphasia, it is not the major decisive factor (Albert et al., 1988; Hickok and Poeppel, 2000).

2.2 The Wernicke-Geschwind Model

Based on the evidence from the described aphasias, Wernicke (1874) formulated an early model of the functional neuroanatomy of language, which was later revised by several authors, most notably Geschwind (1970). It is, therefore, also

commonly known as the "Wernicke-Geschwind Model".

The model proposes that the central elements for language processing are Broca's area, Wernicke's area, the arcuate fasciculus, and the angular gyrus (Mark et al., 2001, p. 625).

Broca's area is the "expressive" area, responsible for production of speech (Broca, 1861). Wernicke's area, the "receptive" area, identifies and analyzes incoming speech (Wernicke, 1874). The arcuate fasciculus is a bundle of axons connecting both areas (Mark et al., 2001, p. 625). The angular gyrus is assumed to be involved in the processing of written language.

How information flows in this system can be illustrated with two tasks (Mark et al., 2001, p. 625): (1) repeating heard speech and (2) reading written text out loud.

(1) Fig. 2.2: Incoming speech sounds, like all sounds, are first processed in the auditory cortex. In order to "understand" these sounds as meaningful linguistic units, they have to be further analyzed in Wernicke's area. Wernicke's area projects via the arcuate fasciculus to Broca's area, which transforms the representations of language into articulatory movement codes. These signals are sent to the motor cortex to (re-)produce the utterance.

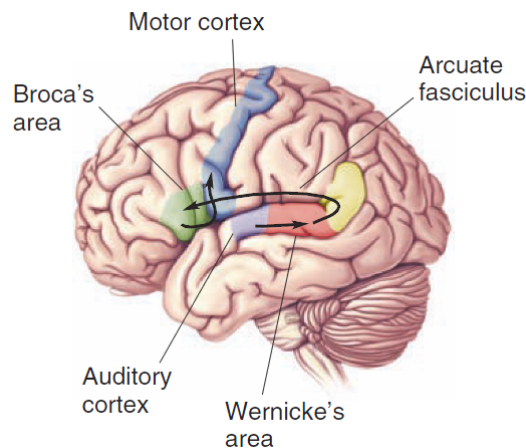


Fig. 2.2: Wernicke-Geschwind model: Repeating heard speech

(2) Fig. 2.3: When reading written language, the information flows from the visual cortex to the angular gyrus which transforms the signal such that it evokes the same activation in Wernicke's area as if the words were spoken (Mark et al., 2001, p. 625). From Wernicke's area onward, the information flow is identical to that in the first task.

2.3 Problems of the Wernicke-Geschwind Model

Whereas the Wernicke-Geschwind model has been highly helpful in guiding research and its interpretations, it has become obsolete as more evidence about the cortical organization of language processes has been gathered (Ward, 2006, p. 250).

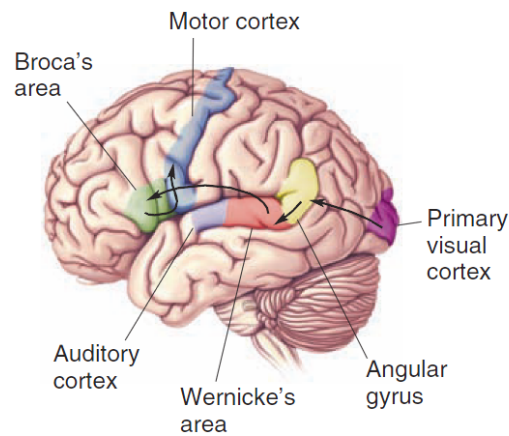


Fig. 2.3: Wernicke-Geschwind model: Saying a written word

The advent of functional brain imaging actuated a variety of promising new possibilities to study the functional neuroanatomy underlying language processes. Notably, Binder et al. (1997) leveraged functional magnetic resonance imaging (fMRI) in an attempt to identify language areas in the human brain. They compared activations for the analysis of speech to activations for the analysis of non-speech sounds. While the language task involved a decision related to the meaning of the utterance, the non-linguistic task required "pitch based decisions about tone sequences" (Binder et al., 1997, p. 353).

The non-linguistic tone task (in comparison to the resting state) induced bilateral activations in cortical regions including the auditory cortex in the STG, premotor cortex, supplementary motor area (SMA), supramarginal gyrus, anterior insula, and anterior cingulate gyrus. The bilateral supramarginal gyrus and planum temporale, and the right posterior MTG responded more strongly in the non-linguistic task than in the language task.

The authors found that all of the regions which were more strongly activated in the language task than in the non-linguistic task were lateralized to the left hemisphere. They identified four large regions related to language: (1) a lateral and ventral temporal lobe region (including the superior temporal sulcus (STS), middle temporal gyrus (MTG), inferior temporal gyrus (ITG), fusiform gyrus and parahippocampal gyrus); (2) a prefrontal region (including the middle and superior frontal gyri, inferior frontal gyrus (IFG) and the anterior cingulate gyrus); (3) the angular gyrus; (4) a "perisplenial region" (including the posterior cingulate, precuneus, and cingulate isthmus).

Hence, whereas the strong left lateralization is consistent with the Wernicke-Geschwind model, other findings by Binder et al. (1997) are less consistent with the classical model as regions other than Wernicke's and Broca's area are activated in the language task: temporal and parietal areas outside Wernicke's area, and frontal areas outside Broca's area. Hence, regions distinct from Wernicke's and Broca's area play a role in speech processing.

There is further evidence that speaks against the Wernicke-Geschwind Model.

For instance, Broca's area is involved in speech perception as well, not only in speech production. As will be shown later in this chapter, it participates in syntactic and semantic processes in speech perception. Thus, Broca's aphasics frequently show comprehension deficits in addition to the production impairments (Caramazza and Zurif, 1976; Ward, 2006, p. 250). The function of Wernicke's area also does not appear to be restricted to perception, but it seems to be involved in production as well. Wernicke himself suggested a participation of Wernicke's area in speech production for Wernicke's aphasics demonstrated disordered speech output (Wernicke, 1874). Moreover, functional imaging studies have revealed left auditory cortex activity during speech production (e.g. Paus et al., 1996). A strict classification into receptive and expressive areas is, therefore, inadequate.

Finally, there is lots of evidence suggesting that different subregions of Wernicke's and Broca's area perform different functions, which would necessitate a subdivision and a more fine-grained functional analysis of both regions.

As we will see, BA 44 is involved in syntactic processes, while the other part of Broca's area, BA 45, additionally to certain syntactic processes also performs semantic processes together with the adjacent BA 47 (see section 2.4.3).

Also in the case of Wernicke's area, part of the posterior STG (pSTG), different portions seem to participate in different mechanisms. Wise et al. (2001, p. 83) assert: "Over time, both the functional and anatomical boundaries of 'Wernicke's area' have become so broad as to be meaningless". The (posterior) STG is divided into (1) the supratemporal plane containing Heschl's gyrus, the planum polare and the planum temporale; and (2) the lateral aspect, which mostly consists of the upper bank of the STS (Kim et al., 2003). Wise et al. (2001) found that the supratemporal plane was activated for both non-speech and speech sounds including the speaker's own voice. The authors propose that its posterior part is related to speech production rather than perception. In contrast, the lateral aspect of the pSTG showed activity in response to speech produced by an external speaker.

In the next section, I will consider the different stages in speech perception in order to establish a comprehensive picture of the current knowledge regarding the function of different brain regions in this process.

2.4 Stages in Speech Perception

All incoming sounds are initially processed in the primary auditory cortex (PAC), be they speech or non-speech sounds (Ward, 2006, p. 224). At some subsequent processing stage, however, the neural systems have to diverge such that speech can be identified and discriminated from non-speech.

The brain has to represent the speech sounds in such a way that their individual linguistic components can be analyzed for the larger structures they form. Traditionally, the smallest linguistic representations of speech are construed in terms of phonemes, abstract categories of speech sounds that make a distinctive difference in the respective language (Wilson and Keil, 2001, p. 787; O'Grady

et al., 1997, p. 60-61).⁴ Different phonemes are concatenated to form syllables, which in turn form words, that are combined to form sentences. Understanding the meaning of complete sentences requires syntactic and semantic processes. Finally, discourse processing enables us to comprehend entire conversations or stories.

In the following, I will cover these individual stages in the speech perception process.⁵

2.4.1 Initial Acoustic-Phonetic Analysis

2.4.1.1 Sound-based Representations of Speech

Hickok and Poeppel (2000) propose that the bilateral STG is the critical cortical substrate for the creation of "sound-based representations of speech"⁶, the initial acoustic-phonetic analysis of the speech signal.

They base their hypothesis on (1) evidence from word deafness and (2) results from functional imaging studies.

Word Deafness

The phenomenon of word deafness provides an important piece of neuroanatomical evidence. In word deafness, speech comprehension is profoundly impaired. Critically, in contrast to Wernicke's aphasia and TSA, word deafness implies profound difficulties in phonemic perception and discrimination (Hickok and Poeppel, 2000).

The major anatomical difference between the two described aphasias (Wernicke's and TSA) and word deafness seems to be lateralization. While Wernicke's aphasia and LSA are associated with damage to the left hemisphere, word deafness occurs almost exclusively when the superior temporal gyrus (STG) is lesioned bilaterally (Buchman et al., 1986).

Functional imaging

During passive perception of speech sounds, bilateral STG activation was found consistently across many studies (Binder et al., 1994; Dhankhar et al., 1997; Mazoyer et al., 1993; Petersen et al., 1988; Price et al., 1996; Schlosser et al.,

⁴ A common way to identify different phonemes of a language is by means of "minimal pairs", pairs of words that differ in only one speech sound (O'Grady et al., 1997, p. 61). The sounds that differ between the two words have to belong to different phonemes since they make a distinctive difference in the language. For instance, "bill" and "pill" are two distinct English words. Hence, [b] and [p] belong to different phonemes.

Note, however, that it is not uncontroversial whether the smallest linguistic components in the speech signal are truly phonemes (Wilson and Keil, 2001, p. 787).

⁵ The main goal of this chapter is to create a concise overview of the brain regions involved in speech perception and their respective functions in this process. There are multiple ways to do this. One may, for instance, choose between a presentation based on neuroanatomy, i.e. the different brain regions, and one that is more functionally-oriented. I chose a functionally-oriented approach as I believe that it provides a more comprehensive picture for the reader. A presentation focusing on each region individually would most likely become too intricate, especially since many regions include subregions which perform different functions in speech processing. On top of that, while such an approach might establish an understanding of the role of each individual region, the reader might not grasp its role within the entire network. After all, we intend to obtain an understanding of the *functional* neural network underlying speech perception.

⁶ Hickok and Poeppel (2004) alternatively use the term "acoustic-phonetic speech codes".

1998; Zatorre et al., 1996). In several of these studies, activity in Broca's area, MTG and the left temporal parietal junction was noted as well, but a lot less consistently. Also, lesions in these regions do not cause the substantial auditory comprehension deficits as lesions in the bilateral STG (Hickok and Poeppel, 2000).⁷

2.4.1.2 Lateralization of Speech Processing

Evidence from aphasias suggests that speech processing is left lateralized. However, if the hypothesis by Hickok and Poeppel (2000) is correct, then the right hemisphere is involved at least up to the processing stage where sound-based representations of speech are established. Hence, the right hemisphere should be able to carry out speech perception processes as well.

Indeed, several studies have shown that when the right hemisphere is isolated⁸, the patient can still understand syntactically simple speech (McGlone, 1984; Wada and Rasmussen, 1960; Zaidel, 1985). Boatman et al. (1995) tested a patient's syllable discrimination ability, while one hemisphere was isolated at a time using intracarotid amobarbital injections. The results indicate that the syllable discrimination ability remained intact no matter which hemisphere was isolated. Thus, both hemispheres seem to be capable of performing syllable discrimination. However, Boatman et al. (1995) found that when a specific area in the left hemisphere was electrically stimulated, the syllable discrimination capacity was impaired. The authors concluded that while both hemispheres are capable of speech perception, the left hemispheric pathway is preferred under normal conditions. When the left hemispheric pathway is unavailable for some reason, the right hemispheric pathway is used instead.

Hickok and Poeppel (2000) also believe that two pathways for speech perception exist. They disagree, however, that the right hemisphere is not important under normal conditions. Instead, they claim that the two pathways "operate in a coordinated fashion, each making a different, but important, contribution to the speech perception process" (Hickok and Poeppel, 2000, p. 134). They base this claim on several studies which demonstrate that whereas the left hemisphere is better than the right at temporal analysis, the right hemisphere is better at spectral analysis (Zatorre, 1997). Robin et al. (1990) performed a lesion study suggesting that the left auditory cortex is important for the perception of temporal information, but not for the perception of spectral information. For the right auditory cortex, the situation is the reverse, i.e. it is significant for the perception of spectral information, but not for the perception of temporal information. Belin et al. (1998) presented sounds with rapid (40 ms) and slow (200 ms) acoustic transitions. For rapid transitions, the response in the left hemisphere was stronger than in the right. For slow transitions, there was an equal response in both hemispheres.

⁷ Further physiological evidence for the significance of the STG in speech perception is given by Creutzfeldt et al. (1989) who illustrated that there are neurons selective to highly specific features of speech, i.e. among others: monosyllabic vs. multisyllabic words, natural vs. distorted / backwards speech, and task-relevant vs. -irrelevant speech. Most of these neurons were found in the middle part of the STG and a few in other temporal areas.

⁸ A hemisphere can be isolated e.g. using carotid amobarbital injections or by investigating split brain patients.

2.4.2 Dorsal and Ventral Streams in Speech Processing

In summary, Hickok and Poeppel (2000) believe that the bilateral STG is responsible for performing the initial acoustic-phonetic analysis of speech.

From this point on, they suggest that two processing streams project further, which are strongly lateralized to the left hemisphere: (1) a **dorsal stream**, which establishes a sensory-motor interface and is involved in sub-lexical processing; and (2) a **ventral stream**, that is responsible for mapping sound onto meaning (Hickok and Poeppel, 2004).

In many ways, this model resembles the neuroanatomical two-streams hypotheses proposed for audition (Rauschecker, 1998) and vision (Goodale and Milner, 1992).

Dorsal stream

The dorsal stream maps sound-based representations of speech onto motor representations for the production of the utterance, which Hickok and Poeppel (2004) call "articulatory-based speech codes". Evidence from functional imaging studies on object naming and articulatory rehearsal tasks suggests that these motor representations of speech are supported by posterior inferior frontal regions (pIF), the frontal operculum (FOP) and a dorsal premotor area (dPM) (Awh et al., 1996; Hickok et al., 2003). The critical cortical structure for mapping sound-based representations of speech onto motor representations might be a region at the boundary of the parietal and temporal lobes in the Sylvian fissure, which Hickok and Poeppel (2004) call "area Spt" (Sylvian-parietal-temporal). The literature provides lots of evidence for the involvement of Broca's area in particular speech perception tasks. Broca's aphasia is associated with syllable discrimination and identification deficits. Moreover, Broca's area is activated in sub-lexical tasks (Burton et al., 2000; Demonet et al., 1992; Fiez et al., 1995; Zatorre et al., 1992), tasks that require explicit attention to phonemic segments.⁹ Hickok and Poeppel (2000), however, argue that such sub-lexical tasks are "fundamentally different from tasks that involve auditory comprehension" (p. 134). They claim that speech perception in everyday life does not require explicit access to segmental information, i.e. information about individual phonemes. Therefore, they hypothesize that sub-lexical tasks involve different neural systems than auditory comprehension tasks. More precisely, sub-lexical tasks recruit the dorsal stream, whereas auditory comprehension implicates the ventral stream (Hickok and Poeppel, 2004).

Ventral stream

The ventral stream maps sound-based representations of speech onto conceptual-semantic representations, i.e. it identifies the conceptual content of speech (Hickok and Poeppel, 2004). Hickok and Poeppel (2004) propose involvement of the superior temporal sulcus (STS) and the posterior inferior temporal lobe (pITL), i.e. the MTG and ITG. As mentioned above, lesions in the pITL often lead to TSA, a post-phonemic comprehension deficit. Thus, they suggest that the function of this region is to map acoustic-phonetic representations of speech onto conceptual-semantic representations. In other words, it collects all phone-

⁹ In a typical study, a subject has to determine whether the last phoneme of two different syllables is the same (Hickok and Poeppel, 2000).

mic, semantic and morpho-syntactic information necessary to interface with the entries in the mental lexicon, the 'lemmas' in psycholinguistic terminology (Levelt, 1999).¹⁰

Their model is summarized in Fig. 2.4. As an example, Fig. 2.5 demonstrates how hearing the word "cat" is processed in this framework.

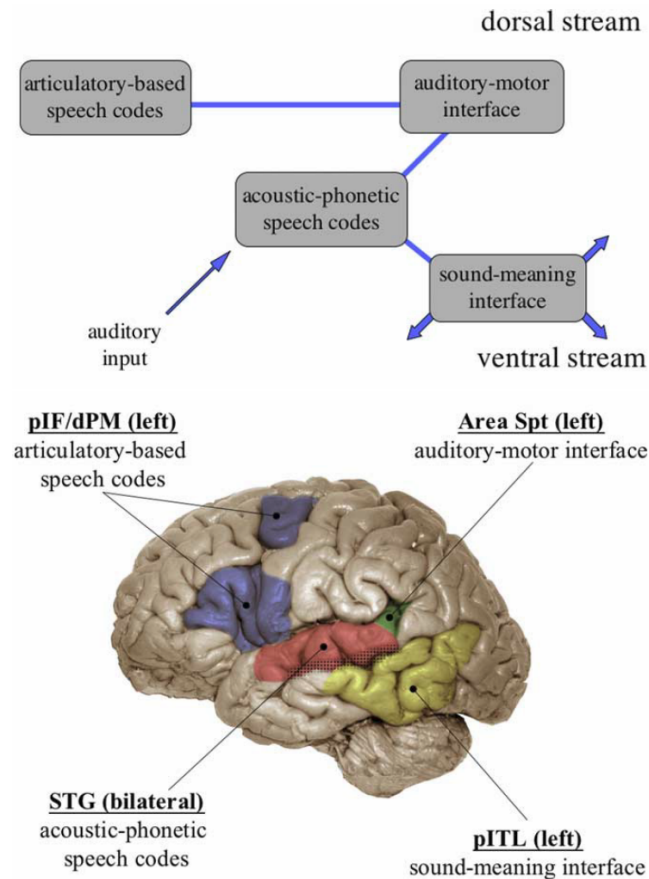


Fig. 2.4: A model of the cortical network underlying speech perception due to Hickok and Poeppel (2004)

2.4.3 Auditory Sentence Processing

After the retrieval of the acoustic-phonetic speech codes (Hickok and Poeppel, 2000), syntactic and semantic processes are required for the comprehension of entire sentences.¹¹ Many researchers have been investigating the precise functions of different regions for sentence-level syntactic and semantic processing.

¹⁰ Hickok and Poeppel (2004) consider Wernicke's aphasia as an impairment of both "auditory speech systems in left STG and sound-meaning mapping systems in left pITL" (p. 94).

¹¹ On top of that, prosodic processes are often necessary to comprehend spoken sentences as well. Frequently, prosody is essential to grasp the intended meaning of the sentence, e.g. when pitch modulations are acting as syntactic markers (see e.g. Meyer et al., 2003). Prosodic

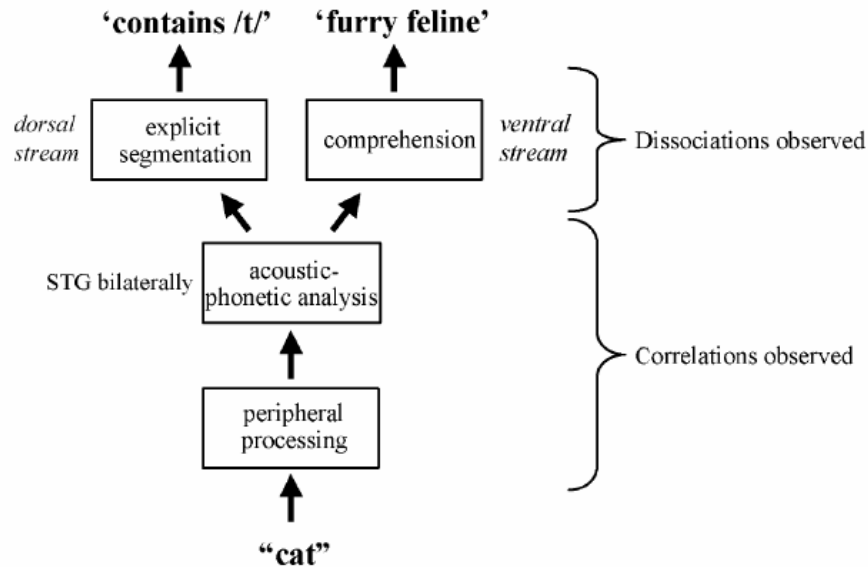


Fig. 2.5: Schema of the processing upon hearing the word "cat" according to the model proposed by Hickok and Poeppel (2004)

In the following, I shall discuss the findings of these studies in detail. It is especially important to note that lots of new evidence has been gathered since the formulation of the dual-stream model by Hickok and Poeppel (2004), which will necessitate an extension and revision of the model.¹²

2.4.3.1 Syntactic Processing

Syntactic processes in sentence processing seem to involve inferior frontal and anterior temporal regions (Friederici, 2002).

It has been suggested that the anterior STG (aSTG) is the critical cortical substrate of phrase structure building processes (Fiebach et al., 2001; Friederici, 2002; Müller et al., 1997). This has been confirmed by imaging studies employing syntactic violations (Meyer et al., 2000). Bornkessel and Schlesewsky (2006) propose that templates of phrase structures are represented in the aSTG/STS. Access to these templates is automatic, which speeds up the phrase structure building process. Friederici (2012) suggests that syntactic relationships between different phrases are established by the aSTG and the FOP.

The left inferior frontal gyrus (IFG), which includes Broca's area, is a highly important area in auditory sentence processing. There is lots of evidence for a functional subdivision within the left IFG. For instance, in a study by Dapretto

processing seems to predominantly involve inferior frontal and superior temporal regions in the right hemisphere (Friederici, 2002 Altmann and Gaskell, 2007, p. 418). This line of research has been omitted from this overview since the right hemisphere is not covered in the experiment conducted for this thesis.

¹² Most of the following studies are collected in Altmann and Gaskell (2007, p. 407-419), Friederici (2002), Friederici (2011) and Friederici (2012).

and Bookheimer (1999), participants had to judge whether two successively presented sentences had the same meaning. Whether the sentences actually meant the same or not depended either on a syntactic or a semantic factor. Syntactic factors elicited an activation of BA 44, while semantic factors activated BA 45/47. Nonetheless, more recent evidence suggests that BA 45 is also important for syntactic processes, albeit its function is not identical to that of BA 44 (Friederici, 2011, p. 1368-1375).

BA 44 seems to play a crucial role when a reanalysis of syntactic structure is necessary, particularly in case of syntactic ambiguity (Altmann and Gaskell, 2007, p. 414-417; Bornkessel et al., 2005; Fiebach et al., 2004). Moreover, Bornkessel et al. (2005) demonstrated that an increase in syntactic complexity also activates BA 44 more strongly.¹³ Several other studies comparing neural activity in response to syntactically simple vs. complex sentences have confirmed a stronger activation of BA 44 for complex sentences (Caplan et al., 1998, 2000; Stromswold et al., 1996). Santi and Grodzinsky (2010) showed that BA 44 is engaged by sentence embedding and syntactic movement, while BA 45 is only activated for movement. This supports the view that BA 44 is "the core region of syntactic complexity" (Friederici, 2011, p. 1371).

However, the factor of syntactic complexity is confounded with working memory factors (Friederici, 2002). Therefore, the results of these and similar experiments have launched a long-lasting debate in the literature: One position argues for a high importance of BA 44 in syntactic processing (Friederici, 2004; Grodzinsky, 2000). According to the alternative view, BA 44 is involved in working memory processes and thus, an increased BA 44 activity for complex sentences is due to higher working memory demands (Caplan et al., 2000; Fiebach et al., 2005; Kaan and Swaab, 2002; Müller et al., 2003). In an attempt to reconcile both views, Friederici (2012) propounds that BA 44 "supports syntactic working memory when sentences are syntactically challenging" (p. 265).

Friederici (2012) proposes that the posterior IFG (Broca's area / BA 44 and 45) also sends top-down predictions to the temporal cortex. For instance, in a sentence where three arguments have been processed, the final verb is expected to take three arguments. Violations of such expectations result in a specific response pattern in the EEG (a biphasic N400-P600)¹⁴ (Friederici et al., 2000).

2.4.3.2 Semantic Processing

Lexical-semantic processes at the word level are supported by the MTG, which is assumed to store conceptual representations (Binder et al., 2009; Lau et al., 2008). The association cortices of both hemispheres appear to be involved as well (Binder et al., 2009).

Semantic processes at the sentence level seem to draw on the anterior temporal lobe, the posterior STG/STS and the angular gyrus (Lau et al., 2008; Obleser et al., 2007a), albeit the exact function of these regions is still controversial (Patterson et al., 2007).

¹³ Syntactic complexity is often defined using the following complexity scale (Altmann and Gaskell, 2007, p.414): conjoined active clauses ("The reporter attacked the senator and admitted the error.") < subject-relative clauses ("The reporter that attacked the senator admitted the error.") < object relative clauses ("The reporter that the senator attacked admitted the error.").

¹⁴ An explanation of ERP responses has been omitted for reasons of brevity.

The anterior temporal lobe has been related to semantic memory (Scott and Wise, 2004). However, other studies have shown that it is not only involved in semantic processes. Vandenberghe et al. (2002) demonstrated that activity in this region reflects both syntactic and semantic factors. Hence, it has been suggested that the anterior temporal lobe integrates different types of information (Scott et al., 2000). Alternatively, it is possible that different subregions of the anterior temporal lobe inhabit different functions in sentence-level speech processing, which would make a more fine-grained analysis necessary (Altmann and Gaskell, 2007, p. 412).

The posterior STG/STS and the angular gyrus are "activated as a function of a word's predictability in sentential context" (Friederici, 2012, p. 264). Activity in the angular gyrus is observed when a word is easily predictable given the linguistic context (Obleser and Kotz, 2010; Obleser et al., 2007a). In contrast, activation in the pSTG/STS reflects a low expectancy of a direct object given its verb (Obleser and Kotz, 2010). These results suggest that the pSTG/STS and the angular gyrus are involved in different functional networks (Friederici, 2012, p. 264).

As mentioned above, BA 45 and 47 play a role in semantic processing. Specifically, these regions are activated whenever lexical-semantic processes are strategically controlled (Friederici, 2012; Newman et al., 2010; Rodd et al., 2005). It has been proposed that BA 45 and 47 access lexical-semantic representations in MTG in a top-down manner (Lau et al., 2008). Note again, however, that activity in the left IFG is associated with an increase in strategic processing demands, while it seems to be much less engaged for sentences that are simple to process (Altmann and Gaskell, 2007, p. 412). This allows for the alternative interpretation that the left IFG primarily supports executive and strategic processes, rather than any language-specific functionality.

2.4.4 A Model of the Functional Neuroanatomy of Auditory Sentence Processing

A comprehensive review of the literature on speech processing up to the stage of sentence comprehension has been provided by Friederici (2012). The author collects recent converging evidence to construct her own model of the functional neuroanatomy of auditory sentence processing, which is depicted in Fig. 2.6. Friederici (2012) supports the dual-stream model introduced above, which was proposed initially by Hickok and Poeppel (2004, revised in 2007) and has been supported by other researchers, e.g. notably Rauschecker and Scott (2009).

After the initial analysis of the sound signal in the bilateral primary auditory cortex (PAC), the initial acoustic-phonetic analysis is conducted: While Hickok and Poeppel (2000, 2004) believe that "sound-based representations of speech" are constructed in the bilateral STG, recent evidence suggests that phonemic processing is performed in the middle portion of the left STG (Friederici, 2012; Leaver and Rauschecker, 2010; Obleser et al., 2007b). From then on, two left-lateralized pathways project further: information flows ventrally via the anterior STG (aSTG) to the frontal cortex and dorsally via pSTG/STS to the premotor cortex (PMC) (Friederici, 2012).

The dorsal pathway comprises two fiber tracts (e.g. Catani et al., 2005; Friederici,

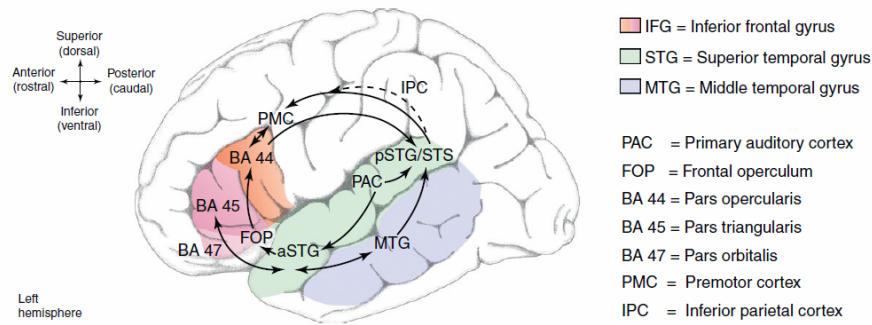


Fig. 2.6: A model of the functional neuroanatomy underlying auditory sentence comprehension by Friederici (2012)

2011): (1) The first tract projects from the temporal cortex to the PMC via the inferior parietal cortex (IPC) and performs a bottom-up sound-to-motor mapping (Hickok and Poeppel, 2004, 2007). (2) The second tract connects the temporal cortex with BA 44 for the processing of syntactically complex sentences (Friederici et al., 2006; Saur et al., 2008; Wilson et al., 2010). BA 44 is assumed to send top-down predictions to the posterior temporal cortex (Friederici, 2012).

The ventral pathway performs sound-to-meaning mapping, as elaborated above (Hickok and Poeppel, 2004, 2007). Projecting from the PAC to the aSTG, larger structures are constructed from their parts (DeWitt and Rauschecker, 2012), "from phonemes to words and phrases" (Friederici, 2012, p. 263). The MTG is proposed to be the most important region supporting lexical-semantic processes (Binder et al., 2009; Lau et al., 2008). BA 45 and 47 perform top-down access to lexical-semantic representations in MTG (Lau et al., 2008). Finally, the aSTG and FOP support the phrase structure building processes (Friederici et al., 2006). The ventral pathway also contains two fiber tracts (Friederici, 2009; Tyler et al., 2011; Weiller et al., 2009): (1) The uncinated fasciculus connects the FOP with the anterior temporal cortex and the temporal pole. (2) The extreme capsule fiber system (ECFS) projects from BA 45 and 47 to the temporal and occipital cortices.

2.4.5 Discourse Processing

What are the neural systems underlying the comprehension of "linguistic units larger than the sentence" (Altmann and Gaskell, 2007, p. 420)? In other words, what are the neural mechanisms that allow us to understand entire conversations or stories?¹⁵

Fletcher et al. (1995) investigated the neuronal activity for stories vs. unrelated sentences and observed that bilateral temporal pole regions, the left pSTG and a region in the posterior cingulate cortex were more strongly activated for stories. When understanding the story required the attribution of a mental state, i.e. when it involved "theory of mind" (TOM) aspects, the left fronto-median area BA 8 was additionally activated (see also Gallagher et al., 2000; Mazoyer

¹⁵ The cited studies are summarized in Altmann and Gaskell (2007, p. 420-421).

et al., 1993).

Ferstl and von Cramon (2001) studied the processing of text coherence by presenting sentence pairs that were coherent or incoherent and cohesive¹⁶ or incohesive. Comparing coherent vs. incoherent sentences, coherent sentences led to a stronger activation of the left posterior cingulate cortex / inferior precuneus, the left fronto-medial wall (BA 9, 10) and the left superior frontal gyrus.

In a subsequent study, Ferstl and von Cramon (2002) presented coherent and unrelated sentence pairs accompanied by an instruction that either encouraged or discouraged the attribution of a mental state, i.e. TOM aspects. Activation of fronto-medial regions (BA 9, 10, 24, 32) was found for all coherent sentence pairs, but for incoherent sentences only when TOM aspects were involved. Hence, the authors conclude that an interaction of coherence and TOM aspects takes place in the fronto-medial wall.

Finally, Xu et al. (2005) compared activations for visually-presented words, sentences and entire narratives. A fronto-temporal network was activated in all conditions. At the sentence level, left inferior frontal and bilateral temporal pole regions were activated as well. At the narrative level, additional activations were found in the bilateral precuneus, fronto-medial cortex (BA 8, 9, 10) and the temporo-parietal-occipital junction.

In summary, there is converging evidence for the involvement of the bilateral temporal pole and the posterior cingulate cortex / precuneus in discourse processing. Fronto-medial regions seem to become engaged when TOM aspects are present. As the study of discourse processing is relatively young, however, much more research is necessary in order to establish a model of the functional neuroanatomy underlying the comprehension of coherent discourse (Altmann and Gaskell, 2007, p. 420).

2.5 Audiovisual Speech Perception

There is lots of evidence supporting the idea that seeing the speaker's lips, i.e. his articulatory movements, can significantly influence and enhance speech processing. The multimodal interaction and integration of visual and auditory information on speech is studied in the field of audiovisual speech perception.

2.5.1 Behavioral Studies

Early studies on the role of vision in speech perception have regarded it as complementary (MacDonald and McGurk, 1978). For instance, Sumbly and Pollack (1954) demonstrated that lip reading improves accuracy in the recognition of spoken words, and Dodd (1977) showed that it enhances speech comprehension in noisy conditions.

Later studies, however, provided strong evidence that visual information about speech is not merely additive, but interacts with the auditory information (MacDonald and McGurk, 1978). McGurk and MacDonald (1976) observed a phenomenon that has become known as the "McGurk effect": When hearing the syllable /ba/, while viewing the lip movements for /ga/, people report hearing

¹⁶ In a cohesive sentence pair, the second sentence began with a connective element like "therefore".

/da/ (81% of preschool children, 98% of adults).¹⁷ This finding illustrates that visual and auditory information about speech interact to derive one integrated percept. Information from both modalities "is combined and synthesized, resulting in the 'auditory' perception of a best fit solution" (MacDonald and McGurk, 1978, p. 254). MacDonald and McGurk (1978) introduced a "manner-place hypothesis" for how visual and auditory information about speech are integrated: The auditory modality determines manner of articulation, while vision determines place of articulation.

2.5.2 Neuroimaging Studies

What is the neurological basis of the multimodal interaction and integration of auditory and visual speech information?

It has been shown that visual information about the speaker's articulatory movements activates the same regions implicated in auditory speech perception.

One of the earliest studies showing that viewing lip movements influences activity in the auditory cortex was conducted by Sams et al. (1991). Using magnetoencephalography (MEG), these authors leveraged the McGurk effect to identify cortical areas involved in the integration of auditory and visual speech information. The participants heard the syllable /pa/, while viewing a face articulating either /pa/ (V = A) or /ka/ (V ≠ A).¹⁸ In the V ≠ A condition, they reported hearing /ka/ or /ta/. The magnetic responses to the V ≠ A condition were similar to the V = A condition in every location except in the supratemporal auditory cortex, i.e. the primary auditory cortex and surrounding regions (Sams et al., 1991). In this area, the deflections began to diverge at 180ms after syllable onset. Thus, as the authors assert, "visual information from the articulatory movements may have an entry into the human auditory cortex" (Sams et al., 1991, p. 143).

This idea has been supported by studies on silent lip reading. Calvert et al. (1997) conducted two fMRI experiments in order to define (1) auditory speech perception regions and (2) regions activated by silent lip reading (vs. watching still lips). They found that there was a significant overlap between regions activated by lip reading and regions activated by heard speech (vs. no auditory stimulus), namely Heschl's gyrus in the left hemisphere (BA 41/42) and the STG in both hemispheres. This suggests that there is a "common code" underlying both auditory and visual information about speech (Altieri, 2010, p. 11). Moreover, the angular gyrus was activated by silent lip reading. This region is traditionally thought to map language-related visual information onto the respective linguistic representations (Calvert et al., 1997; see e.g. Demonet et al., 1992). Calvert et al. (1997) propose that, in this case, the angular gyrus maps information obtained from lip reading onto representations of speech.

Other studies have replicated the activation of auditory cortex by silent lip reading (e.g. Pekkola et al., 2005), albeit its extent is unclear as is its specificity to speech-related stimuli (Campbell, 2008). In general, it can be said that the

¹⁷ A woman was filmed that repeated the respective syllable twice (e.g. ba-ba). This utterance was repeated three times, once every second. The McGurk effect also occurred for other consonant-vowel (CV) syllable combinations: e.g. hearing /pa/ while seeing lips producing /ka/ evokes the perception of /ta/. When the syllables were only presented auditorily, responses were highly accurate (91% pre-school, 99% adults).

¹⁸ The syllable was repeated 800 times, once per second.

middle and posterior STG including the posterior STS (pSTS) show consistent activation across studies on silent lip reading, as well as audiovisual speech (Campbell, 2008, p. 1005).

Neural activity related to lip reading seems to be left-lateralized or bilateral (Calvert and Lewis, 2003; Campbell, 2008; Capek et al., 2004). This stands in striking contrast to the perception of other facial features, e.g. facial expressions or direction of gaze, which typically evoke right-lateralized activations (Campbell, 2008, p. 1005).

Nishitani and Hari (2002) showed that viewing still images which only imply lip movements already engages the mirror-neuron system (MNS). Their study consisted of three conditions: (1) participants watched still pictures of lip movements, (2) imitated them, or (3) executed similar lip movements in a self-paced way. The authors employed MEG to establish the flow of neural activity. As depicted in Fig. 2.7, they found the following sequence of bilateral activations during observation of lip movements: (a) visual cortex, (b) STG, (c) pSTS and inferior parietal lobe (IPL), (d) inferior frontal lobe, and (e) motor cortex. The STS and IPL showed similar activation during observation and imitation, while the inferior frontal lobe and motor cortex were significantly more strongly activated during imitation than in any other condition.¹⁹ According to the authors, these results suggest that "the human mirror-neuron system is activated even by still pictures of lip forms" (Nishitani and Hari, 2002, p. 1217). The MNS, thus, seems to play an important role in the processing of visual speech.

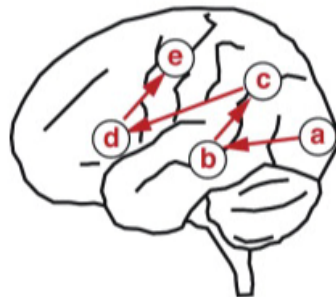


Fig. 2.7: Activation sequence when viewing still images of lip movements according to Nishitani and Hari (2002)

Other studies support an involvement of motor systems. It has been proposed that visual information about articulatory movements may access the "auditory" speech processing systems via motor-based speech representations (Calvert and Campbell, 2003) or that it elicits a "motor plan for the production of the phoneme" (Skipper et al., 2007).

Calvert et al. (2000) propound that audiovisual integration takes place in the left STS. In their study, they compared the responses to an audiovisual stimulus with the responses to the respective unimodal stimuli (audio only and video

¹⁹ All four regions were much more strongly activated during observation of lip movements compared to a neutral lip position. Only the visual cortex showed equally strong activity for both.

only). The audiovisual stimulus was either temporally congruent, i.e. the heard speech was dubbed precisely to the lip movements, or incongruent, i.e. the lip movements did not fit the auditory stimulus. The authors established three criteria to determine whether a region performs integration of auditory and visual speech information. These criteria were set in analogy to the response properties of other multisensory integrative neurons, e.g. in the superior colliculus:

(1) The region should respond to both auditory and visual speech stimuli in isolation.

(2) In the congruent audiovisual condition, the activation should exceed the sum of activations in the two unimodal conditions ("supra-additive activation").

(3) In the incongruent audiovisual condition, the response should be weaker than the sum of the unimodal activations ("sub-additive activation").

The only region fulfilling all three criteria was the left STS. According to the authors, this indicates that the left STS is the site of audiovisual integration of information about speech (Calvert et al., 2000).

The findings by Sekiyama et al. (2003) encourage this supposition. In their fMRI and PET study, these authors used the McGurk effect to determine the neural correlates of cross-modal binding in audiovisual speech perception. The syllables /ba/, /da/ and /ga/ were presented to subjects in three conditions: audio-only (A), video-only (V) and audiovisual (AV).²⁰ The auditory stimuli were additionally varied in signal-to-noise ratio, such that there were two intelligibility conditions (high and low). The behavioral results showed that illusory perceptions occur significantly more often for lower auditory intelligibility. In the audio-only conditions, the same regions were activated in both intelligibility conditions: the STG and STS, angular gyrus and Broca's area (BA 44/45).²¹ In the video-only condition, the major cluster of activity was in MT (V5), which is sensitive to visual stimuli of biological motion. Additional activation was found in Broca's area and the premotor cortex. In the audiovisual conditions, different auditory intelligibility conditions altered the neural activations. For high intelligibility, virtually the same regions were activated as in the audio-only conditions. However, for low intelligibility, the left temporal cortex activity was extended towards MT. This finding corresponds to the behavioral result that a lower auditory intelligibility leads to a stronger visual influence. Finally, the authors directly compared the activations for high and low intelligibility in the AV conditions. They found that there was a significant activation increase in the left pSTS for low intelligibility. Based on these results and previous findings of projections from visual, auditory and somatosensory regions to the STS (Jones and Powell, 1970; Seltzer and Pandya, 1978) and of STS neurons responding to stimuli from different modalities (Desimone and Gross, 1979; Hikosaka et al., 1988)²², they conclude that the left pSTS is involved in cross-modal binding of auditory and visual information about speech.

Campbell (2008) proposes that the pSTS processes the modality-independent dynamic aspects of speech by "abstracting relevant features from both the visual and auditory stream" (p. 1009).

²⁰ All AV stimuli were discrepant combinations of auditory and visual stimuli.

²¹ To assess the brain activity, each condition was compared to a control condition where the subject had to visually identify the still face of a talker.

²² Note, however, that all of the cited studies were performed on monkeys, not humans.

3. EXPERIMENT

3.1 Introduction

In the following experiment, electrocorticography (ECoG) on an epilepsy patient will be leveraged to study the functional neuroanatomy underlying auditory and audiovisual speech perception. ECoG gives the unique opportunity to investigate the neural mechanisms underlying various cognitive tasks invasively, which grants high temporal as well as high spatial resolution. This experiment used the stimulus of a movie to come close to natural conditions of speech perception as they appear in real life, including natural conversational speech that creates a coherent discourse, and seeing the lips of the speaker.

The data will be analyzed for two questions: Which regions in the brain distinguish (i.e. show differential response patterns) between (1) speech vs. non-speech sounds, and which regions differentiate (2) speech with visual information about the speaker's lip movements vs. speech without this information. Finally, the time course of differential activations will be considered.

I hypothesize that the regions sensitive to the contrast in (1) will be those which are implicated in speech processing stages beyond the initial analysis of all types of sounds (speech *and* non-speech) in the primary auditory cortex (PAC), while the regions differentially activated by the contrast in (2) should at least include the presumptive site of audiovisual integration, the (posterior) STS.

3.2 Methods

3.2.1 Participant

The participant was a 19-year old American male and a native speaker of English. He was ambidextrous but wrote exclusively with his right hand. His stay in hospital was for the treatment of medically refractory epilepsy. The experiment was conducted with the participant's (and his parents') written consent.

3.2.2 Apparatus

While the participant was lying in his hospital bed, the electrophysiological activity in his brain was constantly monitored using an invasive imaging method called electrocorticography (ECoG)¹.

¹ The technique is also known as intracranial EEG (iEEG).

ECoG in general

During ECoG, a grid of electrodes is implanted on the exposed surface of the brain in order to record electrophysiological activity intracranially from the cerebral cortex.

Wilder Penfield and Herbert Jasper, neurosurgeons at the Montreal Neurological Institute in the 1950s, pioneered ECoG for the treatment of severe cases of epilepsy (Penfield and Jasper, 1954). Since then, ECoG has been utilized in medically refractory epilepsy to identify the location and extent of so called "epileptogenic zones", areas of the cortex generating epileptic seizures (Kuruvilla and Flink, 2003). These areas are then removed surgically in the procedure of cortical resection.²

In order to perform electrocorticography, a surgeon first removes part of the skull, in a procedure called "craniotomy". The grid of electrodes is then surgically implanted on the cortex surface. The electrodes are placed under the dura mater (subdural). A pre-operative electroencephalogram (EEG) and magnetic resonance imaging (MRI) guide the placement of the electrodes (Schuh and Drury, 1997). The array comprises from 4 to 256 electrodes (Mesgarani and Chang, 2012).

Electrophysiological activity measured by ECoG is composed of local field potentials, generally assumed to reflect postsynaptic potentials (Crone et al., 2001). These potentials are recorded 0.5 to 3mm around the tip of each electrode (Logothetis, 2003).

Both ECoG's temporal resolution of about 5ms and its spatial resolution of 1cm are considered to be high (Mesgarani and Chang, 2012). In non-invasive EEG, electrophysiological activity is measured through the skull. As the conductivity of the bone is low, potentials attenuate very quickly. Thus, as ECoG signals do not have to pass the skull, ECoG's spatial resolution is much higher than that of EEG (Asano et al., 2005). Since a very high spatial resolution is extremely important for the precise localization of the epileptogenic zones, ECoG is currently the "gold standard" measurement of neuronal activity in preparation of surgical treatment of epilepsy (Miller et al., 2007; Sugano et al., 2007).³

² To ascertain whether a patient may come into consideration for resection, a structural lesion of the cortex has to be identified in pre-operative magnetic resonance imaging (MRI). However, even if a brain lesion is found in the pre-operative MRI, ECoG is necessary to determine whether the lesion is in fact the cause of epileptic seizures. Epileptogenic zones may be situated in the area of the lesion, but also around the margin of the lesion or even away from it (Awad et al., 1991).

³ ECoG signals are analyzed for two different types of epilepsy-related ("epileptiform") neurophysiological activity: (1) Neuronal activity during a seizure, so called "ictal" activity. This type of activity is characterized by "fast-wave bursts" often preceded by a strong "leading spike" (Asano et al., 2005). (2) Short spikes of activation between seizures, so called "interictal" epileptiform activity. Asano et al. (2005) found that a "leading spike" was associated with interictal epileptiform activity in all but two out of 42 cases. In case the part of cortex exhibiting this type of activity was removed, 7 out of 8 patients became seizure-free. In contrast, when this part of cortex was not removed, none of the respective patients became seizure-free. Sometimes a region showing epileptiform activity is not removed when it is too close to a functionally important part of cortex (Asano et al., 2005). The major objective of surgery for epilepsy control is to remove the area of cortex causing epileptic seizures, while leaving the functionally important pieces of tissue intact. The extent of cortical resection should be just large enough to effectively control epileptic seizures, not any larger. This is important, for instance, for patients who show epileptiform neuronal activity in their hippocampus. When functionally important parts of the hippocampus are removed, patients often suffer from post-operative memory loss (Kuruvilla and Flink, 2003). Hippocampal ECoG can help to determine exactly which parts of the hippocampus need to be removed to control seizure activity. This

Further advantages of ECoG include that the placement of electrodes is flexible and that recordings can be performed at any stages before and after surgery (Kuruville and Flink, 2003).

ECoG in the present study

In this experiment, a grid of 144 electrodes was implanted subdurally into the cerebral cortex. The electrodes were placed on different regions in the temporal, parietal and occipital lobes in the left hemisphere of the brain. Their sampling rate was 2000 distinct recording samples per second (2000 Hz).

Clinical neurologists from Children’s Hospital Boston marked 14 electrodes (in the temporal and occipital lobes) as ”non-functioning”.

In order to verify these markings, I visually assessed the electrophysiological responses to different speech or sound samples in the movie. In a normally functioning electrode, the curves of electrophysiological responses fluctuate rapidly and they rarely peak above 600 microvolts (mV) or below -600 mV (Fig. 3.1). In contrast, certain electrodes have shown curves with little fluctuations and/or responses as high as 2500 mV. It is highly unlikely (if not impossible) that such responses are valid, i.e. reflect genuine neural activity. Electrodes that show such responses were therefore characterized as ”non-functioning” and excluded from the analyses (e.g. Fig. 3.2)⁴.

I found that for all but one electrode, the electrodes characterized as ”non-functioning” using the described visual detection procedure corresponded to those marked by the neurologists. One electrode (in the middle occipital gyrus) was an exception for it was marked by the neurologists as ”non-functioning” but seemed to show a normal response according to visual evaluation. I decided, nevertheless, to exclude this electrode from the analyses due to the neurologists’ assessment.

In conclusion, excluding the 14 ”non-functioning” out of the 144 electrodes leaves us with 130 functioning electrodes located in the regions listed in Tab. 3.1 which will be used for the subsequent analyses.

can spare functionally important parts, in turn minimizing post-operative memory loss.

⁴ For all plots of electrode responses throughout this thesis, the x-axis shows time in movie frames (24 frames/second) and the y-axis shows the electrophysiological responses in microvolts.

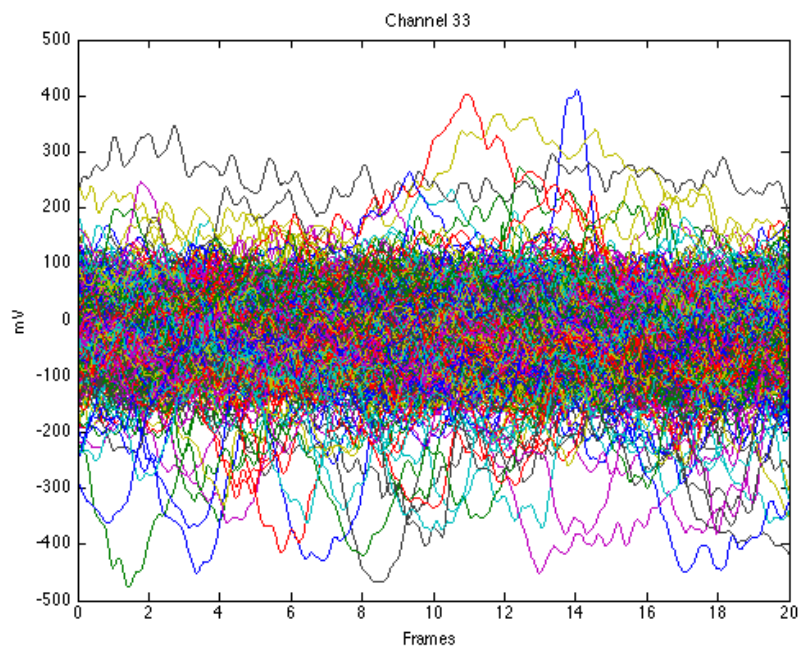


Fig. 3.1: A normally functioning electrode: The response of Channel/Electrode 33 to various sound stimuli in the movie.

Brain region	Functioning electrodes
Inferior temporal gyrus	4
Middle temporal gyrus	10
Superior temporal gyrus (lateral aspect)	4
Superior temporal gyrus (temporal plane)	1
Superior temporal sulcus	2
Supramarginal gyrus	3
Angular gyrus	6
Superior parietal lobe	2
Fusiform gyrus	7
Medial wall	3
Middle occipital gyrus	5
Inferior occipital gyrus and sulcus	1
Superior occipital gyrus	3
Occipital pole	3
Collateral and lingual sulcus	1
Lingual gyrus	8
Parahippocampal gyrus	5
Cuneus	7
Precuneus	35
Cingulate gyrus (posterior-dorsal part)	16
Cingulate isthmus	4
<i>SUM</i>	<i>130</i>

Tab. 3.1: The locations of the 130 functioning electrodes in the left hemisphere.

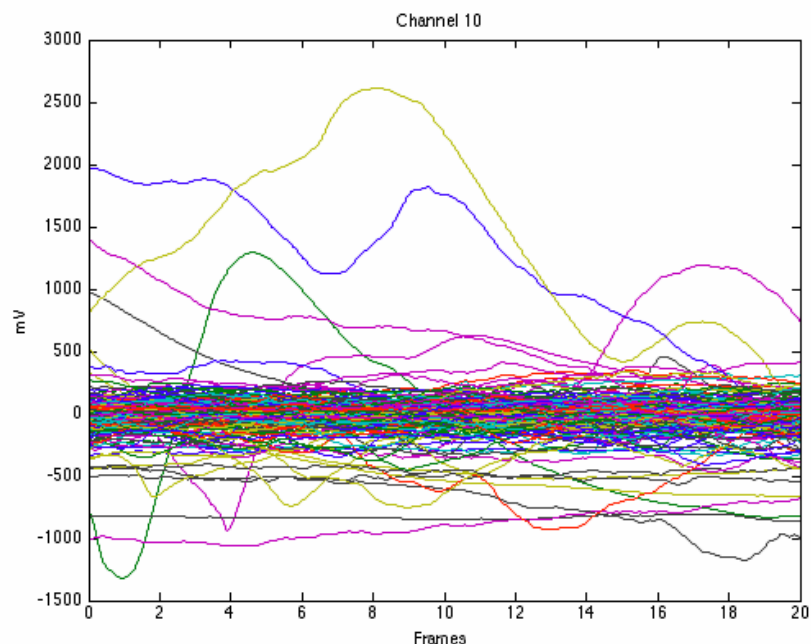


Fig. 3.2: A non-functioning electrode: The response of Electrode 10 to different sound stimuli.

3.2.3 Stimulus Material

The stimulus was the movie "Home Alone 2: Lost in New York", an American Christmas comedy film from 1992. Its frame rate was 24 movie frames per second and it was presented with English audio (without subtitles). The patient watched the movie from a TV screen placed beside the hospital bed in which he was lying.

3.3 Analysis

"Window" of Interest

For each of the following analyses, I chose to consider a temporal "window" of 20 movie frames after speech / sound onset. This length was chosen such that it was long enough to make a response to the respective sound visible (cf. Lau et al., 2008, Friederici, 2002). At the same time, it should not be too long in order to minimize confounds with other stimuli within the same time window. Thus additionally, all sounds were excluded for which a sound of the other category (e.g. speech in case of non-speech and vice versa) occurred within 20 movie frames before or after the respective sound.

Exclusion of Abnormal Responses

Sometimes an electrode may record abnormal responses that deviate very strongly from the mean. It is likely that these responses reflect recording artefacts rather

than genuine electrophysiological activity. For this reason, I excluded all responses whose range (max-min) deviated more than four standard deviations from the median range of all responses.

Baseline Correction

Due to the nature of the movie stimulus, stimuli are present also before the onset of each speech or non-speech sound. As said above, I already controlled for any influence of sounds of the other category (20 movie frames before and after). Still, there are other stimuli appearing before each sound, such as visual stimuli, music etc., which may cause the mean responses to both groups to deviate already before sound onset. While it is of course infeasible to control for every imaginable confounding factor, it is feasible to normalize the responses such that they are as close to each other as possible at the time of sound onset. To this end, a typical baseline correction procedure was applied (cf. Bledowski et al., 2004; Gruber and Müller, 2005; Mulert et al., 2004; Spencer et al., 1999). A period of 5 movie frames, i.e. approximately 200ms, was defined as the baseline period.⁵ Then, for both categories, the mean of the response within the baseline period was taken and subtracted from the response at each sampling point. The result of this procedure is that the mean response of both groups within the baseline period is 0. The procedure thereby corrects for any systematic deviations before sound onset without changing the response patterns to the actual sound stimulus.⁶

Statistical Test

For each electrode, the response to all sounds in one category was compared to the response to all sounds in the other. Hence, a paired sample t-test was taken over the two groups at each sampling point (2000 sampling points are in a second).

The type I error (α), i.e. rejecting the true null hypothesis that there is no difference between the responses to the two categories, was chosen to be 0.05.⁷

Threshold for Period of Significance

However, a significant difference computed in this manner should only "count" as significant if the difference lasts over a longer period of time. Thus, we additionally define a numeric threshold for the period of time the significance has to be maintained. For now, we will choose a threshold of 60 sampling points. Later, we will compute a threshold using a statistical measure for the final analysis when controlling for various visual and auditory factors.

⁵ In the literature, around 200ms before stimulus onset is typically taken as the baseline period (e.g. Bledowski et al., 2004; Mulert et al., 2004). 5 movie frames correspond to 208.3ms ($5 \cdot 1000 / 24$).

⁶ Both the exclusion of abnormal responses and the baseline correction procedure were done in the algorithm presented in section 5.3.4.

⁷ For a MATLAB algorithm performing the statistical test, see section 5.3.1.

3.4 Results

3.4.1 Speech vs. Non-Speech Sounds

In the first series of analyses, I compared the electrophysiological responses for speech stimuli to the responses for non-speech sound stimuli, in an attempt to determine which areas of the brain distinguish speech from non-speech sounds.

The onset time of each sound sample was manually determined at the accuracy of one movie frame. For the speech samples, each onset was the movie frame corresponding to the onset of a sentence heard in the movie.⁸ For the non-speech sound samples, the onset corresponded to the start of a salient environmental sound. These included bangs, squeaks, fireworks, explosions, gunshots, glass shattering, wood cracking, water splashing, telephone ringing, knocking, foot-steps, car sounds, emergency sirens, and animal sounds such as the twittering of birds.⁹

I decided to compare an equal number of responses for speech and for non-speech sound samples. There were 1299 samples of speech and 152 of non-speech sounds. Averaging over a very large number of responses yields a flat curve, i.e. it eradicates any pattern that may underlie the data. Furthermore, a substantial difference between the number of samples in each category might actually produce a difference in the responses when a genuine response pattern is compared to a flat curve (e.g. Fig. 3.3). Hence, to avoid these averaging effects, I randomly drew the same number of speech samples as non-speech sound samples. The responses to 152 non-speech sound samples and 152 speech samples were thus compared.

55 electrodes showed a significant difference. However, the differential response to speech vs. non-speech could be due to other factors than a differentiation between speech and non-speech alone. Certain auditory and/or visual stimuli in the movie might correlate with either speech or non-speech. In the following, I will attempt to control for the most striking confounding factors.

3.4.1.1 Control for Faces

The most salient confounding variable in the visual domain is a face shown on the screen. It seems likely that a face is seen much more often when someone

⁸ Note that the exact temporal onset of a sound is not always completely clear. Especially in the case of speech, it is often debatable. For instance, in a word like "house" (/haus/), the /h/ sound is barely audible individually; only when the /a/ sound is produced, do we truly hear the /h/ as well. Therefore, I decided to align the onset of speech to the onset of voicing. The temporal onset of non-speech sounds is much clearer.

⁹ Non-speech human vocal sounds, like laughing, screaming etc. were not included in this analysis as their processing could be different, i.e. it might involve other neural mechanisms and/or pathways. Research on voice processing suggests that the human voice in general, even when not producing speech, carries "important affective and identity information" (Belin et al., 2004). The bilateral STS has been shown to be selective to voices (Belin et al., 2000), while the right anterior STS seems to be important for the identification of a particular speaker and, thus, the representation of unique, individual voices (Belin and Zatorre, 2003; Warren et al., 2006). In conclusion, it is likely that the processing of non-speech vocal sounds differs from that of non-speech non-vocal sounds, and hence these types of sounds have been excluded from the analysis.

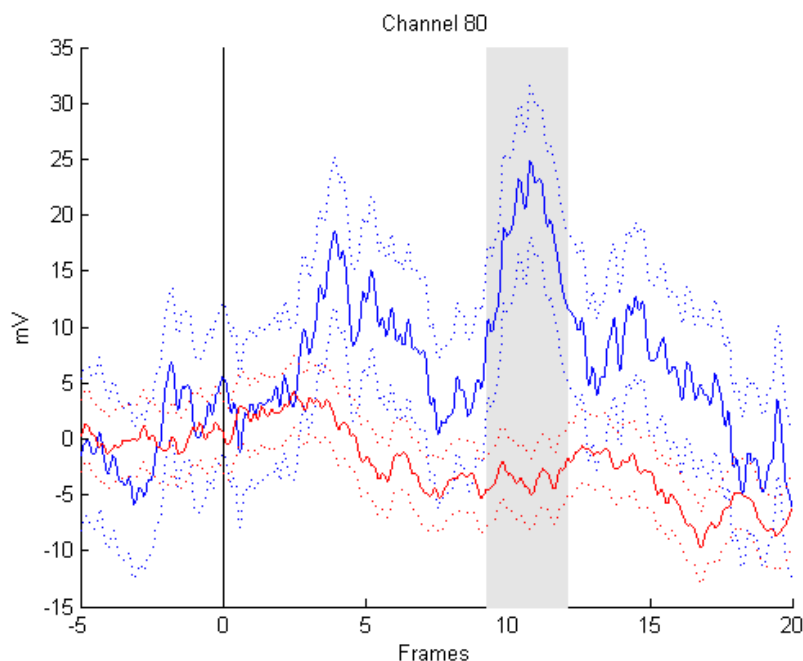


Fig. 3.3: The mean response of Channel/Electrode 80 to speech (red) and non-speech sounds (blue). The dotted lines represent the respective standard deviation and the grey area shows a period of significant difference between the responses.

speaks compared to when a salient non-speech sound appears. Therefore, in the following, all speech and non-speech sound samples that occurred together with a face shown on the screen will be excluded from the analysis. Only the responses for speech and non-speech sound samples that occurred in absence of a face will be compared. Excluding all samples with faces leaves us with 80 speech and 56 non-speech sound samples.

3.4.1.2 Controls for Auditory Features

In the auditory domain, two possibly confounding factors come to mind: loudness and pitch / spectral frequency. It is possible that differences in the responses for speech and non-speech may be due to systematic differences in loudness and pitch in the two groups. For instance, it seems that many salient non-speech sounds, especially explosions, gun shots and the like are much louder than a typical speech sample. For these reasons, the 80 speech and 56 non-speech sound samples that we obtained by excluding all samples occurring when a face was shown on the screen, will additionally be controlled for loudness and spectral frequency.¹⁰

Loudness Control

In order to control for the factor of loudness, I considered the distribution of

¹⁰ See section 5.3.2 for an algorithm that executes both controls.

peak magnitudes for both groups (converted to decibels (dB)) as shown in Fig. 3.4.

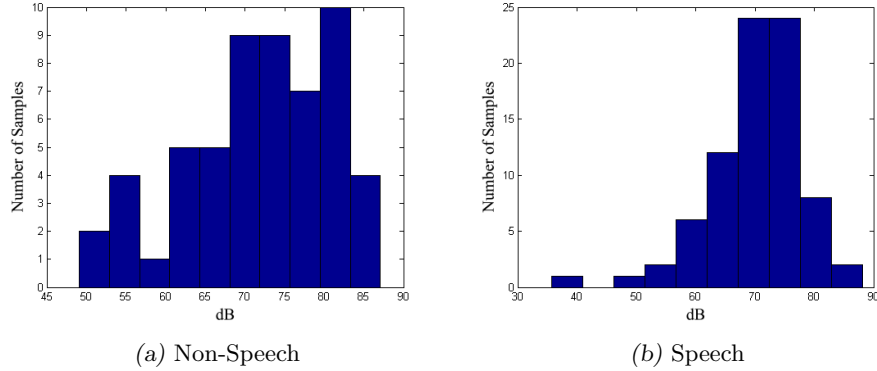


Fig. 3.4: Distributions of peak magnitudes (loudness) for each sample in decibels (dB).

We find that there are indeed much more non-speech sound samples that are very loud (right end of the graph) than speech samples. Hence, I defined a loudness threshold such that the mean loudness of each group was as close to the other as possible. A loudness threshold of 80dB was appropriate to achieve this goal. All samples that were louder than this threshold were excluded from the analysis. This process reduced the number of non-speech samples from 56 to 42 and the number of speech samples from 80 to 75. After the application of the threshold, the mean loudness of non-speech was 36.73dB and the mean loudness of speech was 36.81dB.

Spectral Frequency Control

I also intended to control for possible differences in pitch / spectral frequency. Instead of randomly drawing an equal number of samples for both groups as before, I tried to find those speech samples that were most similar in their frequency spectrum to the non-speech sound samples. To this end, the discrete Fourier transform (DFT) was computed for each speech and non-speech sound sample. For each of the 42 non-speech sound samples, the speech sample with the closest spectral frequency was chosen and included in the analysis. Concretely, I computed the area between curves between the DFT of the respective non-speech sound and the DFT of the speech sample. The speech sample with the minimal area between curves was chosen. This process was repeated for each non-speech sound.

All in all, the responses for 42 non-speech sound samples and 42 speech samples were compared.

Threshold for Period of Significance

Let us reconsider the threshold we defined for how long a significant difference period has to last so that we are convinced that the electrode indeed shows a systematic difference in its response to speech vs. non-speech sounds. Before, we rather arbitrarily chose this threshold to be 60 sampling points. For this final analysis, we want to find a threshold justified by a statistical measure.

In line with the type I error (α) being 0.05 for the t-test, this threshold was

computed such that a comparison of two random groups yielded a significant difference at most 5% of the time. The idea behind this measure is that ideally, only the two groups of speech and non-speech should produce distinct responses of the electrode. Hence, a random assignment of samples to two groups should only very rarely also produce a differential response. An equal number (42) of samples was randomly drawn from all 42 speech and 42 non-speech sound samples controlled for faces, loudness and spectral frequency and assigned to either a first or a second group. For each electrode, the threshold was computed such that a significant difference was found in at most 5 out of 100 runs.¹¹ The mean of the thresholds of all electrodes was 114 sampling points, which was taken as the threshold for a period of significance for the following analyses.¹²

Results

25 electrodes in the following regions demonstrated a significant difference sustained over a period of 114 sampling points or more¹³: MTG (2), STG (planum temporale (1)), supramarginal gyrus (1), angular gyrus (1), superior parietal lobe (2), fusiform gyrus (4), medial wall (1), occipital pole (1), lingual gyrus (3), cuneus (1), precuneus (5), and cingulate isthmus (3).

For the plots of the electrophysiological responses in these electrodes, see section 5.1.1 of the appendix.¹⁴

Fig. 3.5 shows the time course of activations for speech vs. non-speech. Each "differentiation time" is the time at which the first significant difference period¹⁵ began in the respective electrode. Thus, in layman's terms, it is the time at which the particular region started to clearly differentiate speech from non-speech sounds. The coloring represents temporal categories, which merely serve the purpose of data visualization.¹⁶

3.4.2 Speech with Seen Lips vs. Speech without Seen Lips

In the second analysis, I will contrast the activity to speech with visual information on the speaker's lip movements vs. speech without that information. There were 894 samples with seen lips of the speaker and 233 without lips.

Note that in 133 out of the 233 samples without lip movements, faces are still seen, i.e. faces of characters other than the respective speaker. This already controls for the possibly confounding factor of seen faces to a great extent for it is not the case that faces only co-occur with samples from the 'speech with

¹¹ For each run, 42 samples were randomly assigned to each of the two groups again. The responses were normalized according to the baseline correction method described above.

¹² The algorithm that puts the described procedure into execution is depicted in section 5.3.3

¹³ The number in parentheses after each region signifies the number of electrodes showing a significant difference in that region.

¹⁴ The algorithm shown in section 5.3.4 was used to determine these results. It plots the responses of each electrode and computes all significant difference periods (as shown under 'Time Course' below).

¹⁵ There can be multiple significant difference periods if there is a period in between them where the difference was not significant anymore (according to the t-test with $\alpha = 0.05$).

¹⁶ As these temporal categories are only introduced to be able to mentally conceptualize the time course of differential activations more easily, they are chosen "by hand" and not using any statistical methods.

very early (<150ms); early (150-250ms); in between (250-350ms); late (350-450ms); very late (>450ms)

Regions	Differentiation times (start time of first significant difference period for each electrode)				
Superior parietal lobe	2.392f = 99.7ms	7.288f = 303.7ms			
Angular gyrus	3.316f = 138.2ms				
Cingulate isthmus	3.856f = 160.7ms	4.696f = 195.7ms	5.26f = 219.2ms		
Precuneus	4.408f = 183.7ms	4.528f = 188.7ms	5.176f = 215.7ms	5.788f = 241.2ms	11.92f = 496.7ms
Cuneus	4.636f = 193.2ms				
Supramarginal gyrus	5.032f = 209.7ms				
STG (planum temporale)	5.08f = 211.7ms				
Fusiform gyrus	5.152f = 214.7ms	15.16f = 631.7ms	15.712f = 654.7ms	18.556f = 773.2ms	
Medial wall	5.32f = 221.7ms				
Occipital pole	6.316f = 263.2ms				
MTG	8.848f = 368.7ms	18.4f = 766.7ms			
Lingual gyrus	10.216f = 425.7ms	12.292f = 512.2ms	18.628f = 776.2ms		

Fig. 3.5: The time course of differential activations for speech vs. non-speech.

seen lips' category, but a face is also shown in over half of the samples from the 'speech without seen lips' category.

3.4.2.1 Controls for Auditory Features

Loudness Control

For the control of loudness, let us again consider the distributions of peak loudness for the two categories (Fig. 3.6). As one can see, the distributions are highly similar. Also the mean loudness is very close: 37.04dB for speech without seen lips and 36.77dB for speech with seen lips. Therefore, it is unnecessary to undertake some sort of additional control for loudness.

Spectral Frequency Control

Using the same procedure as above, for each of the 233 samples without seen lips, that sample with seen lip movements was drawn out of the 894 which was the closest in spectral frequency, i.e. the one that yielded the minimal area between curves of the DFT of the two sounds. Hence, the responses to 233 samples with lips vs. 233 samples without lips were compared.

Results

A significant difference period of at least 144 sampling points was identified in 22 electrodes in the following locations: inferior temporal gyrus (2), MTG (5), STG (lateral aspect (2), planum temporale (1)), STS (1), angular gyrus (2),

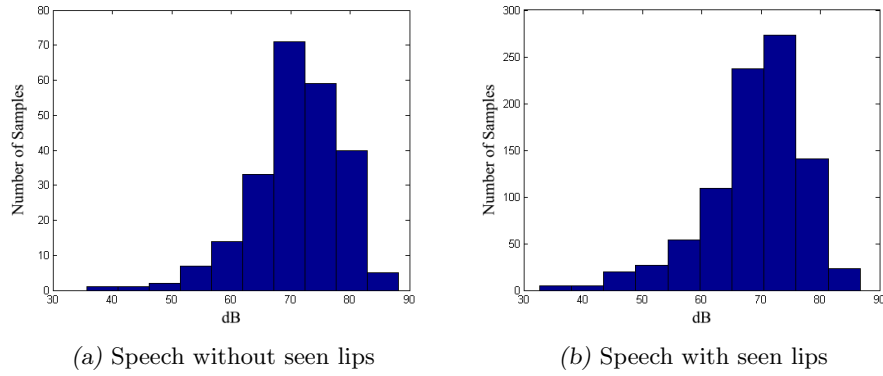


Fig. 3.6: Distributions of peak magnitudes (loudness) for each sample in decibels (dB).

fusiform gyrus (1), middle occipital gyrus (1), superior occipital gyrus (1), pre-cuneus (3), cingulate gyrus (posterior-dorsal part (1), isthmus (2)). Again, the plots of these electrodes can be found in the appendix (section 5.1.2).

The time course of differential responses to speech with seen lips vs. speech without seen lips is depicted in Fig. 3.7.

very early (<200ms); early (200-300ms); in between (300-450ms); late (450-550ms); very late (>550ms)

Regions	Differentiation times (start time of first significant difference period for each electrode)				
MTG	3.928f = 163.7ms	10.888f = 453.7ms	13.912f = 579.7ms	17.152f = 714.7ms	17.44f = 726.7ms
STG (planum temporale)	5.428f = 226.2ms				
Precuneus	5.908f = 246.2ms	9.064f = 377.7ms	9.328f = 388.7ms		
STG (lateral aspect)	6.16f = 256.7ms	15.508f = 646.2ms			
Cingulate isthmus	8.776 = 365.7ms	9.544f = 397.7ms	11.536f = 480.7ms		
Angular gyrus	10.072f = 419.7ms				
Cingulate gyrus (posterior-dorsal part)	10.18f = 424.2ms				
Superior occipital gyrus	11.164f = 465.2ms				
Inferior temporal gyrus	11.404f = 475.2ms	15.088f = 628.7ms			
Middle occipital gyrus	12.616f = 525.7ms				
Superior temporal sulcus	15.088f = 628.7ms				
Fusiform gyrus	15.172f = 632.2ms				

Fig. 3.7: The time course of differential activations for speech with seen lips vs. speech without seen lips.

3.5 Discussion

3.5.1 Speech vs. Non-Speech Sounds

As I hypothesized, regions distinguishing speech from non-speech sounds are mostly regions involved in speech processing stages beyond the initial auditory analysis in the primary auditory cortex (PAC), which treats all kinds of sounds equally (Ward, 2006, p. 224). Nonetheless, we also find a few exceptions of regions differentially activated for speech vs. non-speech that do not belong to the typical speech processing areas.

For several brain regions which show a significant difference, it should not be surprising that they distinguish speech from non-speech sounds considering previous findings in the speech processing literature (see chapter 2).

To a large extent, I managed to replicate the results of a very similar study by Binder et al. (1997) who compared the neural activity for a task on heard speech to the activity for a task on non-speech sounds. Just like these authors, I also found differential activations in the left STG, MTG, supramarginal gyrus, fusiform gyrus, angular gyrus, cingulate isthmus and the medial wall.

The STG has been implicated as one of the major language areas in the brain for a long time (e.g. as part of Wernicke's area (Wernicke, 1874)). It has been proposed that the STG is the site where "sound-based representations of speech" are created (Hickok and Poeppel, 2000, 2004), that its left middle portion executes the initial "acoustic-phonological analysis and the processing of phonemes" (Friederici, 2012) and that the projection from the PAC to the anterior STG constructs larger structures from their parts, "moving from phonemes to words and phrases" (Friederici, 2012, p. 263; cf. DeWitt and Rauschecker, 2012).

I observed differential responses in the planum temporale (PT), part of the posterior STG (pSTG) (Westbury et al., 1999). Studies by Binder et al. (1996, 1997) have shown that the PT responds more strongly to non-speech sounds than speech during active listening. They, therefore, propose that the PT is "involved in early auditory" (Binder et al., 1996, p. 1239) rather than language-specific processing. Similarly, the supramarginal gyrus was more strongly activated in the task based on non-speech sounds than in the speech task (Binder et al., 1997).

However, as described in section 2.4.3.2, the pSTG/STS and the angular gyrus are related to semantic processes at the sentence level. They respond differently depending on how predictable a word is given the linguistic context (Friederici, 2012, p. 264): While the angular gyrus is activated for easily predictable words, the pSTG/STS shows activity for words that are rather unexpected (Obleser and Kotz, 2010; Obleser et al., 2007a). Hence, in contrast to the proposal by Binder et al. (1996), the PT might be involved in language-specific processes.

The MTG is associated with semantic processing at the word level (see section 2.4.3.2). It is thought to hold the conceptual representations of individual lexical entries (Lau et al., 2008). Thus, it is part of the ventral stream mapping sound-based representations of speech onto semantic representations (Hickok and Poeppel, 2004).

Differential responses were also found in the cingulate isthmus and precuneus. Both of these regions are thought to play a role in discourse processing (see

section 2.4.5). Activation in these regions should be expected for a narrative movie, after all, tells a story. Hence, sentences are highly related; even entire conversations have to be related semantically to other conversations. As the dorsal part of the posterior cingulate gyrus is not differentially activated, it is possible that only the ventral part, i.e. the cingulate isthmus, but not the dorsal part, is involved in discourse processing.

At this point, my findings diverge from those of Binder et al. (1997). A few regions were differentially activated in my experiment which were not activated in the study by Binder et al. (1997) and which are not typically implicated in speech perception processes, including the lingual gyrus, superior parietal lobe, cuneus, and occipital pole.

The lingual gyrus is typically implicated in word processing during reading (Horwitz et al., 1998; Mechelli et al., 2000; Price et al., 1996; Price, 2000; Shaywitz et al., 2002; Vitacco et al., 2002). However, several studies on the functional neuroanatomy of speech processing observe activity in the lingual gyrus. Rodd et al. (2005) found bilateral lingual gyrus activation during the perception of low-ambiguity speech¹⁷ vs. a baseline noise condition. Zekveld et al. (2006) compared activations for different signal-to-noise ratios (SNR) of speech. The bilateral lingual gyrus was activated¹⁸ as soon as speech became intelligible, i.e. for intermediate and high SNRs, but not for low SNRs, where speech was unintelligible. Still, activity in the lingual gyrus during speech perception is often disregarded as reflecting processes that are rather related to reading than to the processing of speech. For instance, Zekveld et al. (2006) argue: "activation observed in this region most likely reflects processes not directly associated with the intelligibility of the speech, such as visualization of the potential graphic structure of the words". Certainly, it is highly unlikely that lingual gyrus activation during the perception of speech in a movie is due to a visualization of words. Thus, the results of my experiment show that we should consider that the lingual gyrus does play a role in language-related processes specific to speech perception.¹⁹

Activity in the superior parietal lobe (SPL) is rarely observed during speech perception. An exception is a study by Zatorre et al. (1992) who found that

¹⁷ A low-ambiguity sentence was e.g. "Her secrets were written in her diary", whereas a high-ambiguity sentence included several ambiguous words, e.g. "The *shell* was *fired* towards the *tank*".

¹⁸ Activity for speech was compared to activity for only noise.

¹⁹ Similar ideas also seem to hold for the fusiform gyrus (FG). Like the lingual gyrus, the FG (or rather a subregion within it) is most often implicated in visual word processing during reading rather than a language-specific process in speech perception (Binder et al., 2006; Dehaene et al., 2002; Devlin et al., 2006; McCandliss et al., 2003). However, its function in reading is often related to phonological processing, e.g. for speaking written words out loud (Cao et al., 2008; Dietz et al., 2005). Moreover, FG activity is frequently observed during the perception of auditory speech (Buckner et al., 2000; Chee et al., 1999; Demonet et al., 1992, 1994; Giraud and Price, 2001; Perani et al., 1998; Wise et al., 2000). It was also engaged in the semantic task on speech in the experiment by Binder et al. (1997) and it distinguishes speech from non-speech sounds in my experiment. Nevertheless, several researchers believe that the language-sensitive region in the fusiform gyrus is a "visual word form area" (e.g. Dehaene et al., 2002; McCandliss et al., 2003). Dehaene et al. (2002) hypothesize that the occasional recruitment of the FG during speech perception is due to the formation of a "mental image of the written form of spoken words" or of "mental images appropriate to the semantic contents of spoken words" (p. 324).

the left SPL is much more strongly activated during a phonetic judgment task²⁰ than during passive speech perception. This suggests that the left SPL is involved in active, attentional aspects of perceiving speech. It is certainly possible that such processes are engaged during the perception of speech in a movie.

The cuneus and occipital pole belong to the traditional early visual areas. Now, it is not impossible for visual cortex to be selectively activated by speech. Giraud et al. (2001) demonstrated that in cochlear implant (CI) patients, early visual cortex is recruited during speech processing in the absence of visual stimulation. In a subsequent study, early visual cortex as well as the left posterior ITG and SPL were more strongly activated when listening to native language speech vs. speech from an unknown language, in both CI patients and normal-hearing subjects (Giraud and Truy, 2002). Thus, the authors assert that "auditory-to-visual cross-modal effects can also be recruited under natural hearing conditions" (p. 1562). Admittedly, however, it is also possible that visual regions show a differential response due to a confounding variable in the visual domain. This may include salient low-level visual changes such as a cut in the movie.²¹

3.5.2 Speech with Seen Lips vs. Speech without Seen Lips

One region has been singled out in the literature as the site of integration of auditory and visual information on speech: the STS (see section 2.5.2, Calvert et al., 2000; Campbell, 2008; Sekiyama et al., 2003). It is, therefore, as expected that the STS responds differently for speech with seen lips than for speech without seen lips. It is also clear why the lateral aspect of the STG showed differential responses for it consists mostly of the upper bank of the STS (see section 2.3, Kim et al., 2003).

The planum temporale (PT) of the STG is differentially activated as well, which presents further evidence against the idea by Binder et al. (1996) that PT involvement is restricted to early auditory processing. The PT is part of the pSTG, which is consistently activated for both audiovisual speech and silent lip reading (Campbell, 2008, p. 1005).

Another region that distinguishes speech with seen articulatory movements of the speaker from speech without that information is the angular gyrus. Activity in this region during silent lip reading has been observed by Calvert et al. (1997). Based on the classical idea prevalent also in the Wernicke-Geschwind model (see section 2.2) that the angular gyrus maps linguistic information from the visual domain onto representations of language (Demonet et al., 1992), Calvert et al. (1997) hypothesize that it maps visual information about the speaker's lip movements onto speech representations. My results are consistent with this idea.

For several other regions, it is less clear why they respond differently to speech with seen lips than to speech without seen lips, including the MTG²², ITG, precuneus and posterior cingulate, fusiform gyrus, and the superior and middle occipital gyrus.

²⁰ Subjects had to respond in case the last consonant sound of two syllables was the same.

²¹ It would be possible to control for this factor in the future by removing all samples containing a cut, albeit this has the danger of reducing the number of samples too greatly.

²² One piece of evidence for the involvement of MTG in the processing of audiovisual speech is provided by Callan et al. (2004) who illustrated that the MTG is more strongly activated when the speaker's face and lips are clearly visible than when the visibility of the face is reduced (using a spatial low-pass filter).

Many of these regions are involved in speech processing. As we saw in section 2.5.2, visual information on the speaker's lip movements has an entry into the auditory cortex and essentially engages the "auditory" speech perception regions. Already the observation of still pictures of lip movements suffices to activate these regions. Hence, it has been suggested that some "common code" underlies auditory as well as visual information on speech (Altieri, 2010, p. 11). Now, the result from this experiment that many regions thought to be involved in speech perception are differentially activated by speech with seen lip movements vs. speech without seen lips adds new perspectives to these ideas.

Firstly, we find regions involved in the earliest stages to regions involved in the latest stages of speech processing to be affected by visual information on the speaker's lip movements: From the STG (i.a. initial acoustic-phonetic analysis) over the MTG (semantic processing) to the precuneus and posterior cingulate (discourse processing), all of these regions are sensitive to the contrast between speech with seen lips vs. speech without seen lips. This strongly supports the view that visual information on the speaker's articulatory movements significantly influences speech processing.

Secondly, while there may be a "common code" (Altieri, 2010, p. 11) underlying both auditory and visual speech information, the finding that the response of many speech processing regions differs for speech with seen lips vs. speech without seen lips gives reason to suppose that these regions still treat (i.e. process / analyze) visual information on the speaker's lip movements differently than auditory speech information.

3.5.3 Time Course

It is possible that the temporal sequence of activations reflects the flow of information from one area to the next. It is also possible, however, that these regions do not communicate at all and instead receive signals from a third source. The question which of these options is truly the case cannot be answered based on these data alone, but it would be necessary to perform a connectivity analysis (e.g. using the MATLAB toolbox for Granger causal connectivity analysis (Seth, 2010)).²³

For several regions, there were strong differences in the differentiation times across different electrodes in that region. Considering speech vs. non-speech, severe discrepancies were found in the superior parietal lobe, precuneus, MTG, lingual gyrus and fusiform gyrus. For speech with seen lips vs. speech without seen lips, differentiation times of different electrodes differed immensely in the MTG²⁴, ITG, and the lateral aspect of the STG. Other regions which displayed disparities, although not as intensely, were the precuneus and cingulate isthmus.

²³ While a connectivity analysis is beyond the scope of this thesis, it does constitute an exciting opportunity for future research.

²⁴ The MTG showed very early differentiation for speech with seen lips vs. speech without seen lips, even earlier than for speech vs. non-speech. This may seem counter-intuitive at first. However, looking at the MTG electrode which demonstrated a very early significant difference for speech with seen lips vs. speech without seen lips, we notice that this electrode also shows a very early differentiation for speech vs. non-speech, but this difference was not sustained long enough (not at least 114 sampling points) to count as significant. So in general, it seems that (at least part of) the MTG distinguishes between speech and non-speech, and also speech with lips vs. speech without lips very quickly.

These findings strongly prefigure that the same region is part of different functional networks or pathways in speech perception. Hence, a functional subdivision of the affected regions seems necessary. As elaborated in chapter 2, in the last decades, we have been witnessing a trend towards a more fine-grained anatomical subdivision of a number of language-related areas due to functional differences (Friederici, 2012): To name a few, Broca's area is functionally separated into BA 44 and 45 for they participate in different functions in syntactic processes and BA 45 is also engaged in semantic processes; and the left STG is split up into a middle portion involved in the processing of phonemes, an anterior portion (aSTG) partaking in the ventral stream mapping sound to meaning, and a posterior portion (pSTG) as part of the dorsal stream performing a sound-to-motor mapping. Now, the temporal data from this experiment indicates that we may also need to functionally separate different subregions within other areas such as the precuneus, lingual gyrus or MTG.

4. CONCLUSION

The experiment presented in this thesis provides intriguing new viewpoints on the processing of auditory and audiovisual speech in the left hemisphere of the brain.

As we saw in chapter 2, speech processing consists of several stages: All sounds, speech or non-speech, are incipiently processed in the primary auditory cortex (PAC). Afterwards, speech is identified and discriminated from non-speech, such that sound-based representations of speech are created. These representations are then analyzed for their linguistic components. The smallest components, typically represented as phonemes, are combined to form syllables, syllables create words and words build sentences. The comprehension of whole sentences necessitates syntactic and semantic processes.¹ Finally, understanding interrelated sentences, i.e. coherent conversations or stories as in a narrative movie like "Home Alone 2" requires processing of discourse.

The first analysis of the experiment comparing the responses to speech vs. non-speech revealed, as hypothesized, differential activations of regions involved in all speech processing stages beyond the initial auditory analysis in the PAC: from early phonemic processing in the STG, to semantic processes at the word-level in the MTG and at the sentence level in the angular gyrus and pSTG, to discourse processing involving the precuneus and cingulate isthmus. Regions which distinguished speech from non-speech that are not typically implicated in speech processing included the SPL, the lingual gyrus and early visual cortex (cuneus and occipital pole). The SPL may partake in attentional aspects of speech perception. The lingual gyrus is often viewed as an area whose language-related function is restricted to reading, even though activity in this region is observed frequently during speech perception. The findings of this experiment, however, suggest that it may play a significant role in speech processing. It is puzzling that early visual cortex is recruited. While this could reflect a confounding visual factor, it is also possible that the visual cortex is engaged by a genuine language-related process such as the processing of native language speech.

The section on audiovisual speech perception (section 2.5) illustrated that audiovisual speech essentially engages the same network of regions as auditory-only speech; even visual speech alone can activate these areas. This was taken as evidence that there is a "common code" (Altieri, 2010, p. 11) underlying both auditory and visual speech information. The left (posterior) STS was identified as the cortical substrate of the integration of auditory and visual information on speech.

¹ And, as noted in section 2.4.3, prosodic processes are often required as well.

Now, the second analysis of my experiment looked at the differential responses to speech with seen lips vs. speech without seen lips. The presumed site of audiovisual integration, the STS, was differentially engaged as expected. The majority of the other regions showing a significantly different response were speech processing regions. This supports that visual information on the lip movements of the speaker engages the "auditory" speech processing regions and significantly alters speech processing in these regions. So whereas a "common code" for speech from both modalities might exist, the visual lip reading information is still processed differently from auditory speech.

The time course of differential activations illustrates striking divergences between "differentiation times" of different electrodes in the same region. It is, thus, likely that different subregions of the same region participate in different functional networks in speech processing. Therefore, functional subdivisions like the ones that regions such as Broca's area and the left STG have gone through in the last decades may be necessary for other regions as well, including the MTG or precuneus.

5. APPENDIX

5.1 Electrode Plots

This section shows the plots of all electrodes with a significant difference over a period of at least 144 sampling points.

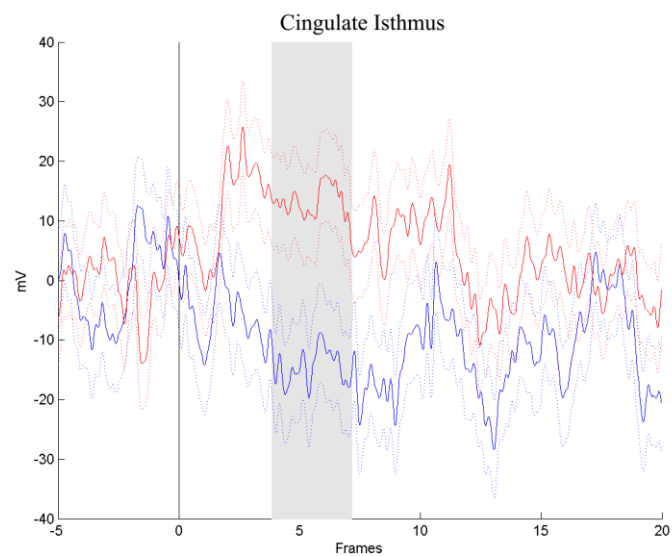
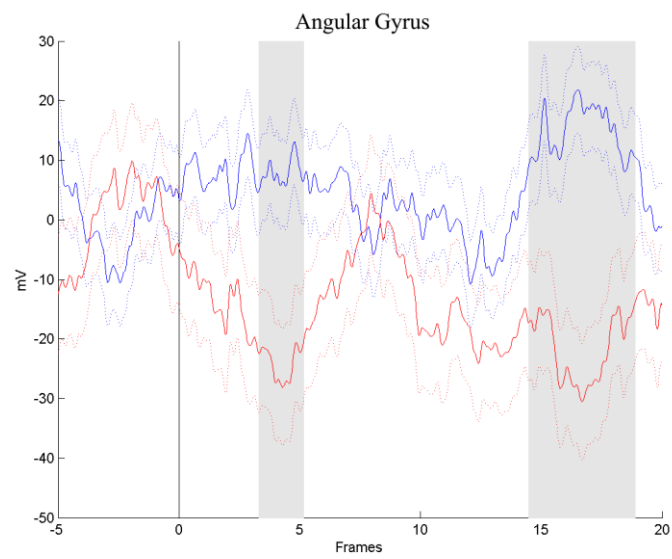
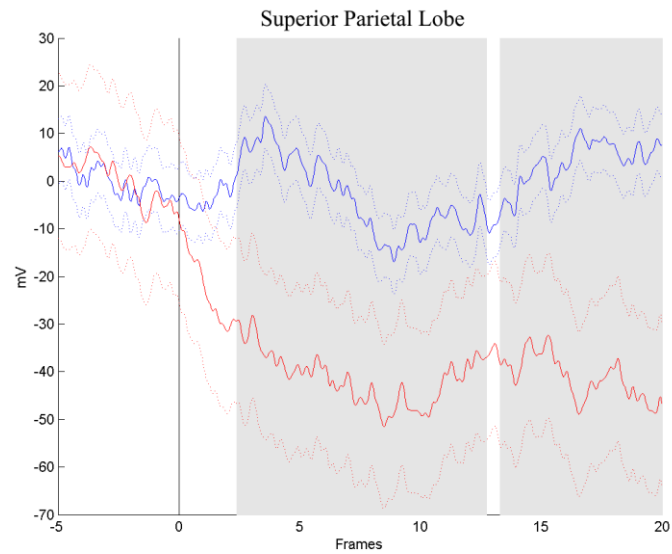
In each plot, the x-axis represents the time in movie frames (24 per second), while the electrophysiological activity given in microvolts is seen on the y-axis. The vertical black line at $x = 0$ illustrates sound onset. After this point, the time window of 20 movie frames which was considered for the analyses is depicted, as is the baseline period¹ of 5 frames before sound onset. Solid lines show the mean of the responses to all samples of the respective group. Dotted lines display the standard deviation. Finally, grey areas highlight the periods of a significant difference between the responses to the two-groups according to the t-test with $\alpha = 0.05$. Such a period had to last at least 144 sampling points (2000 per second) to "count" as significant (see section 3.4.1 for how this threshold was determined). Periods of a shorter duration are not displayed.

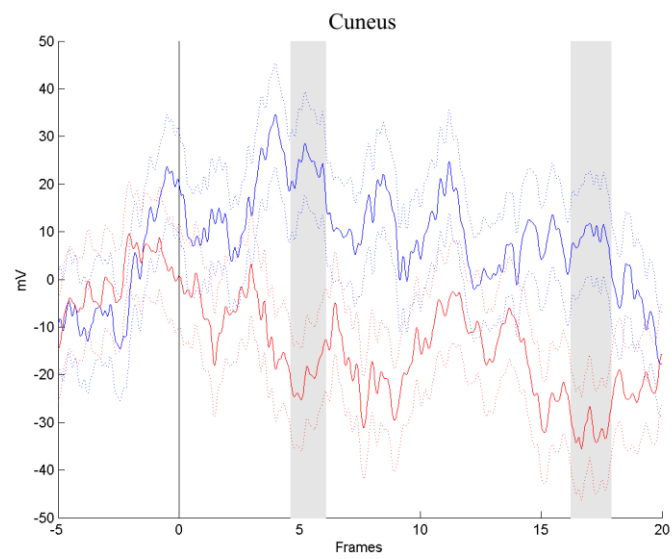
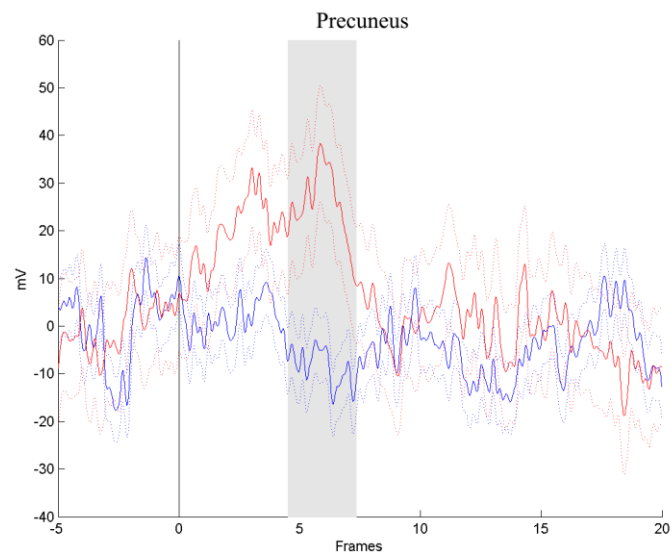
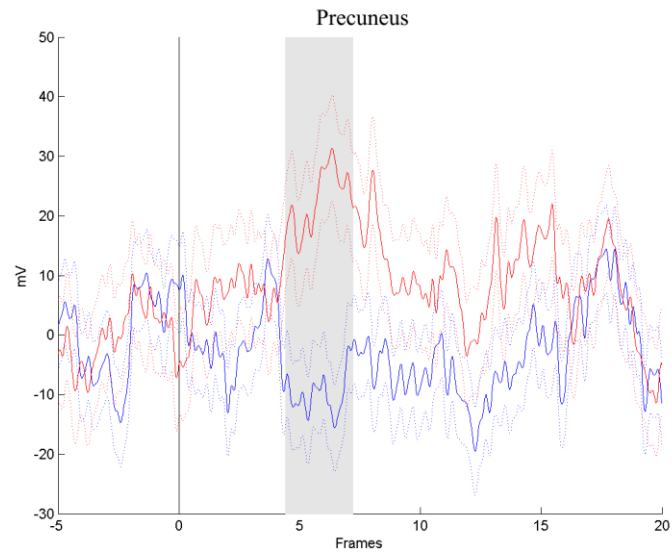
The plots are arranged in the temporal order of the differentiation times (see time course).

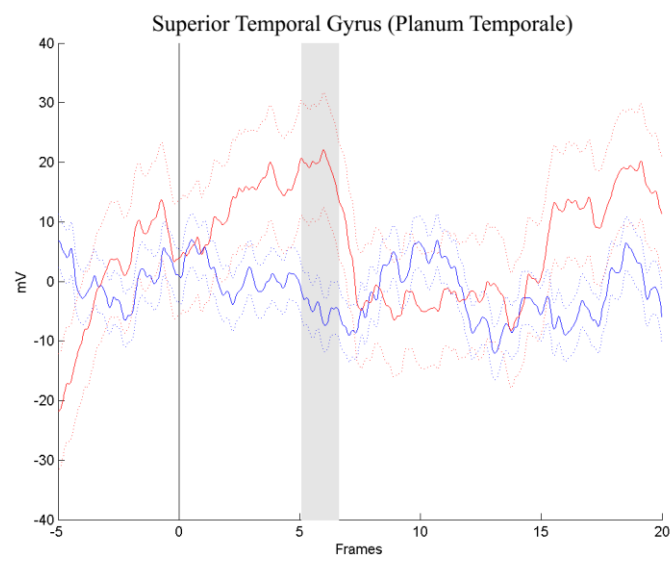
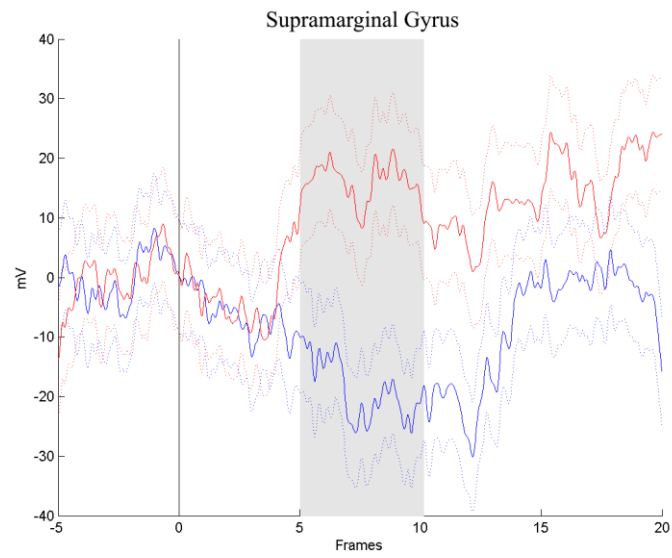
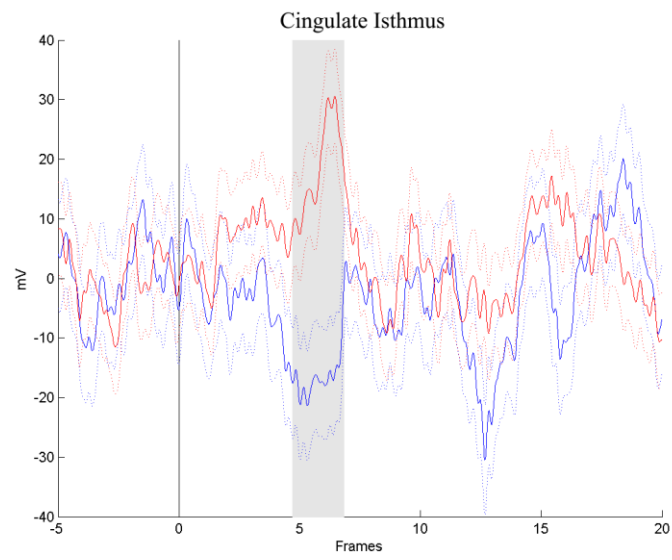
¹ For the baseline correction procedure, see section 3.3.

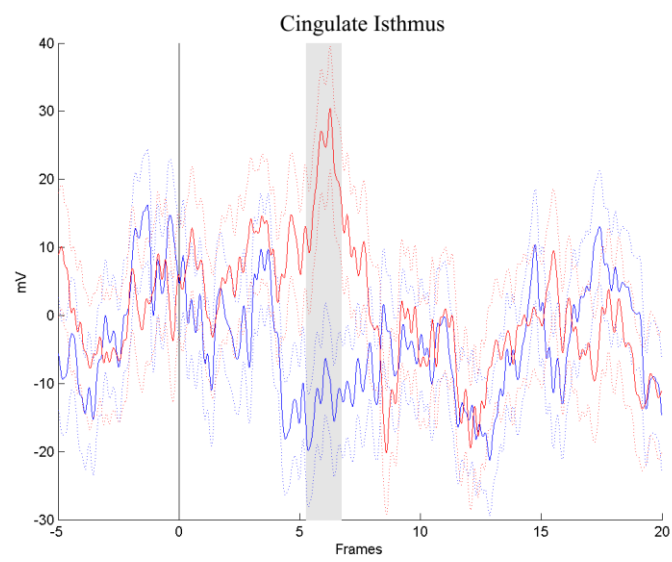
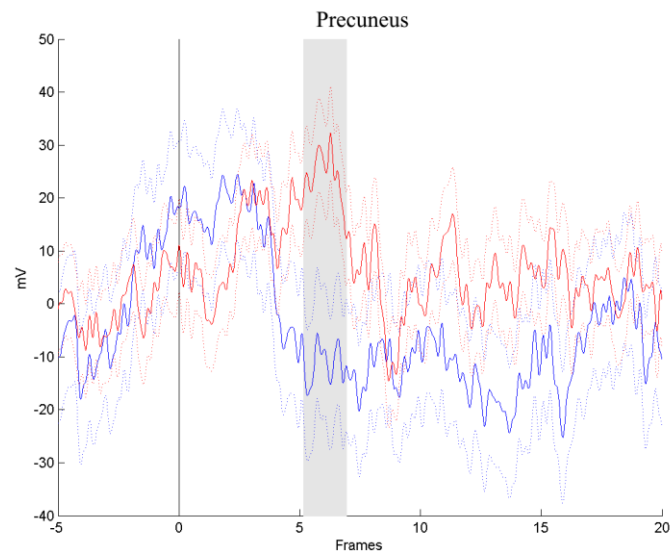
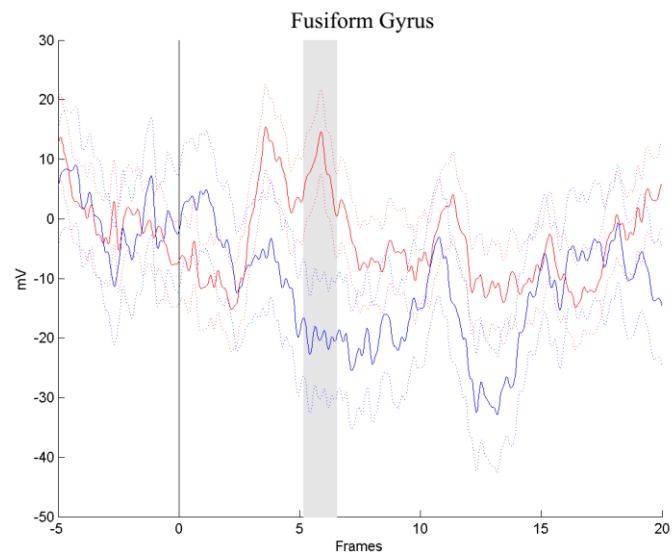
5.1.1 Speech vs. Non-Speech Sounds

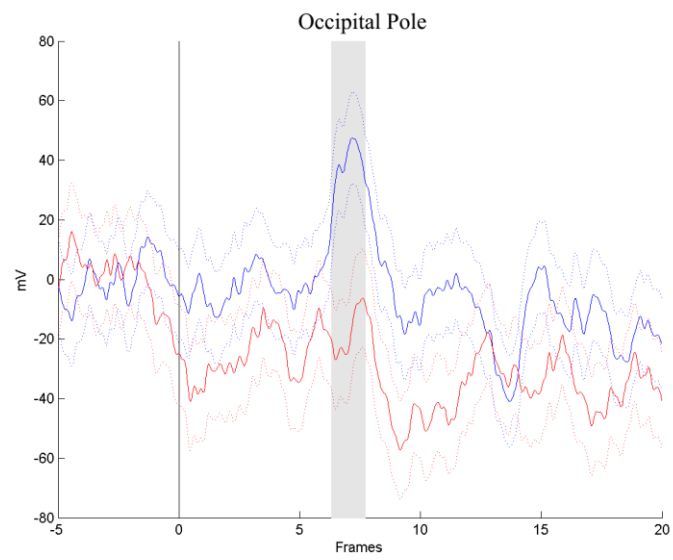
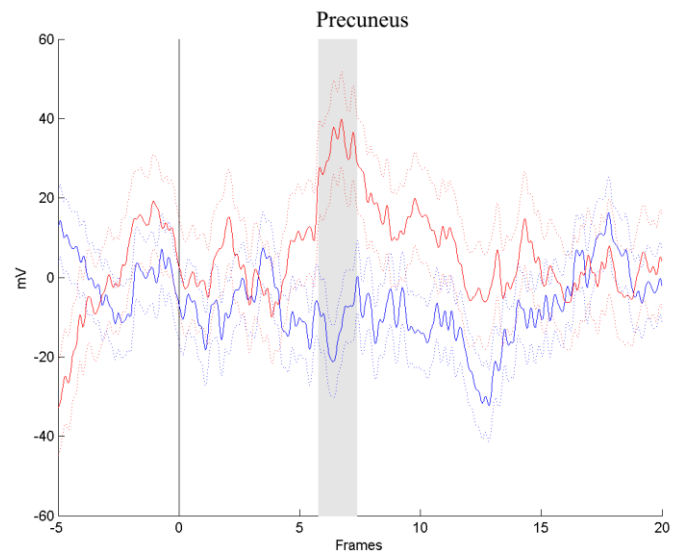
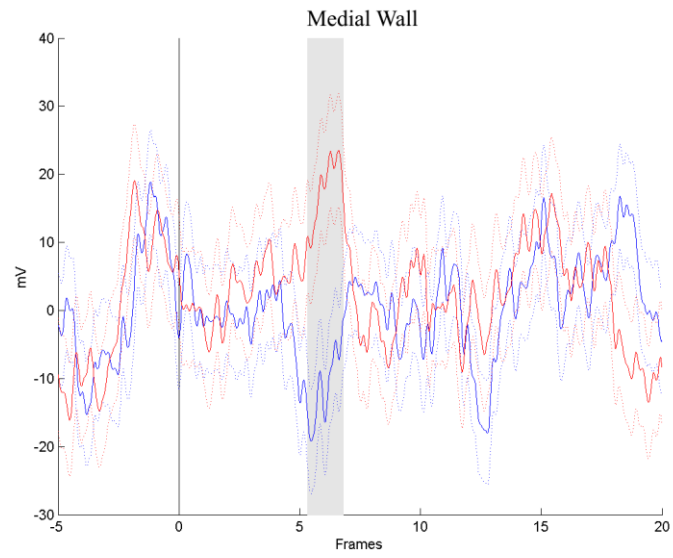
red: speech; blue: non-speech

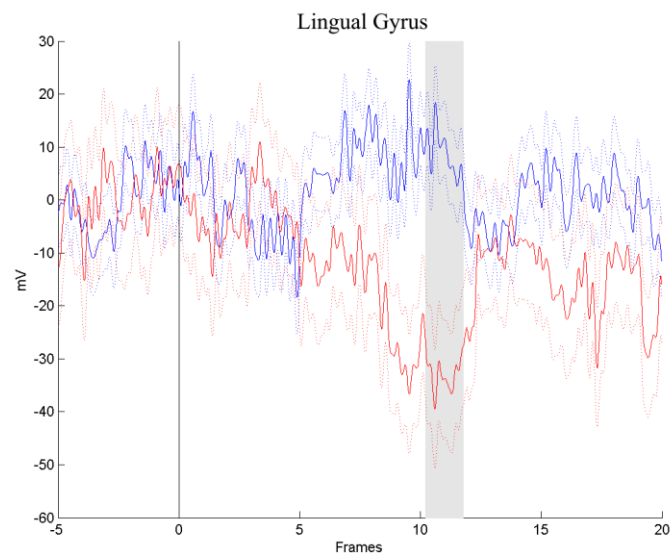
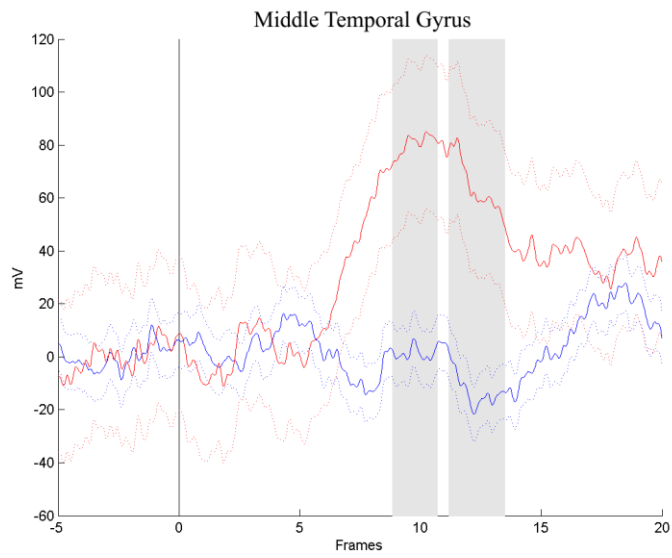
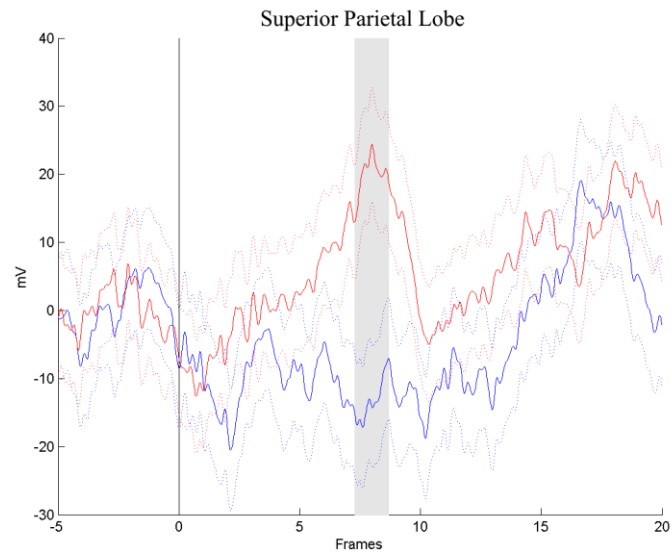


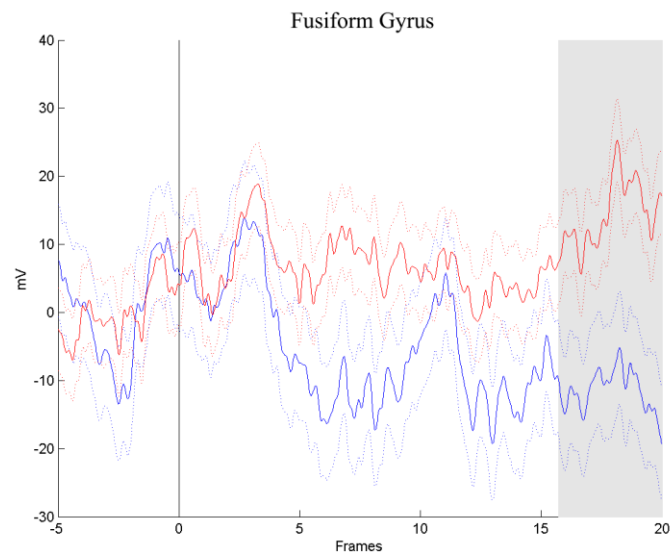
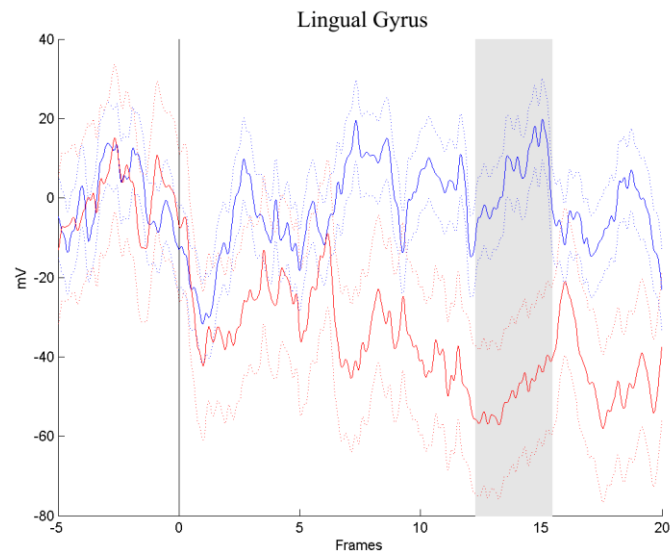
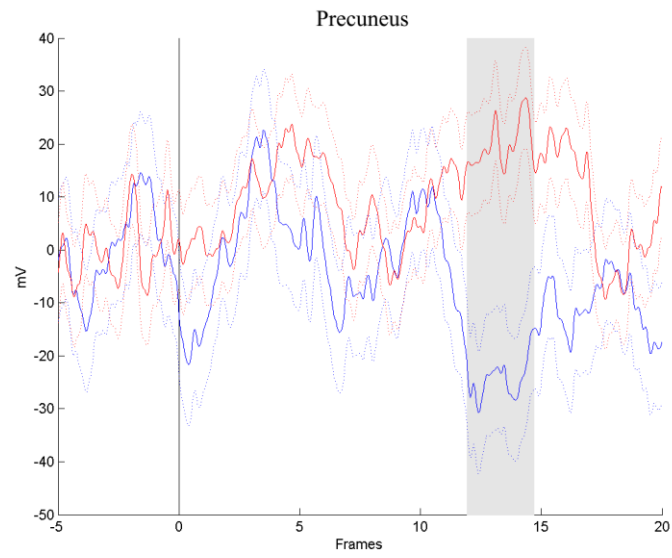


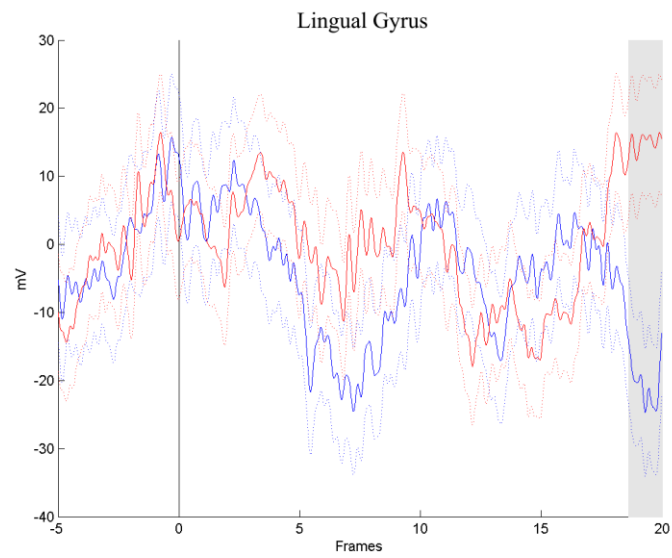
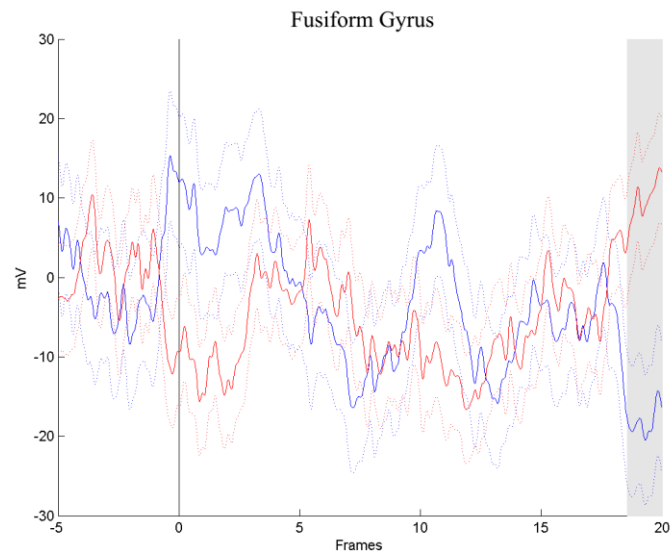
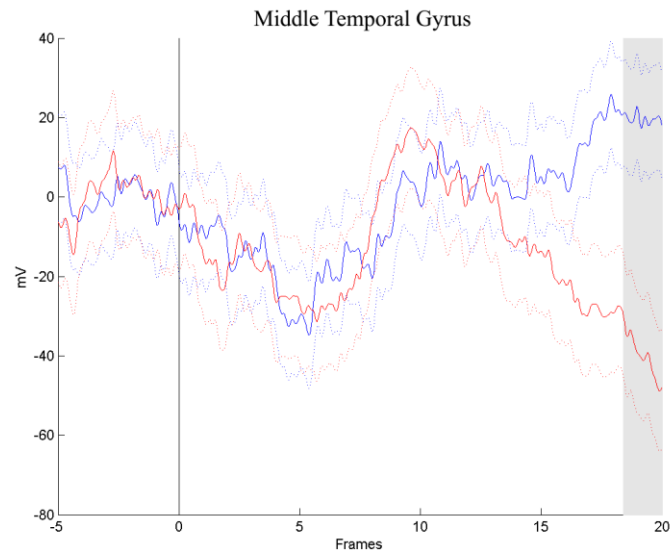






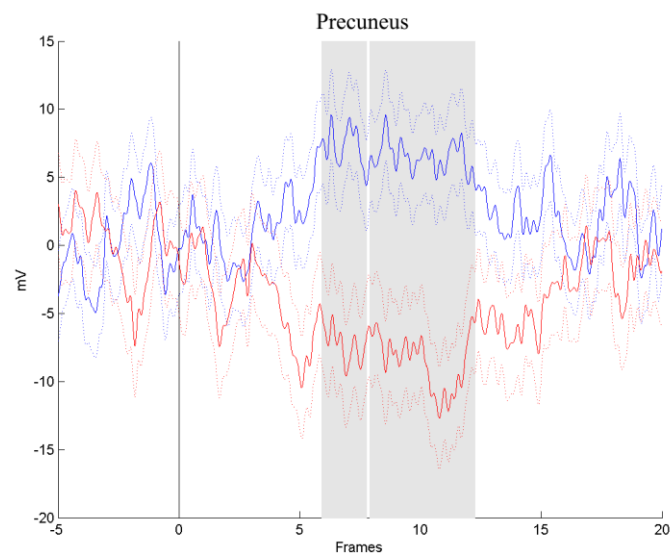
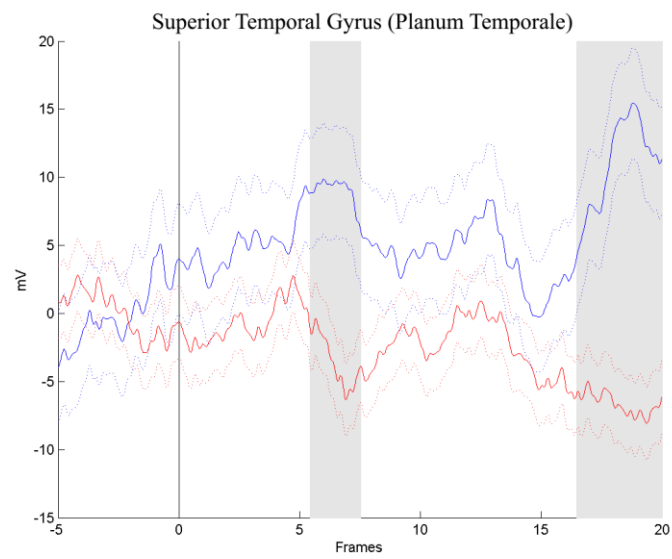
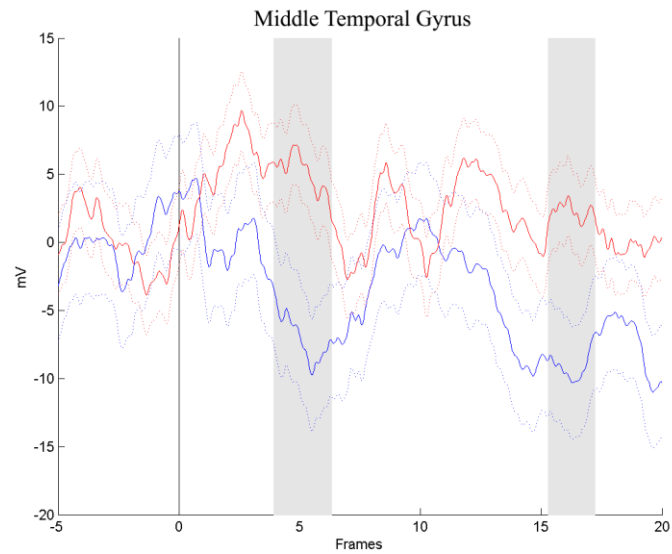


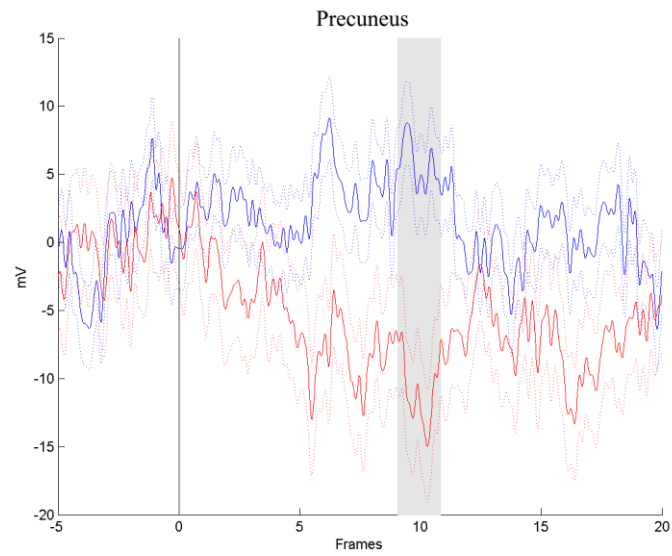
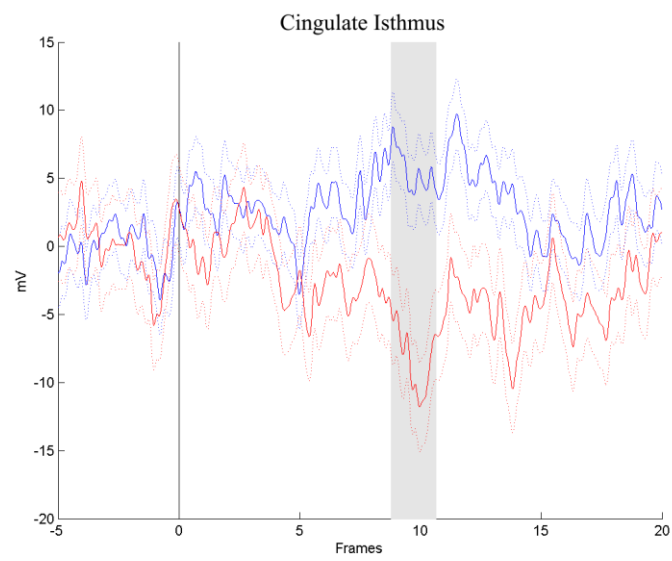
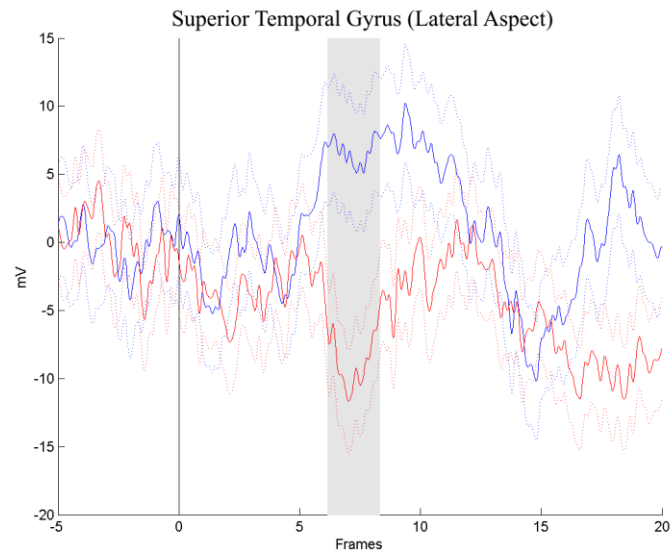


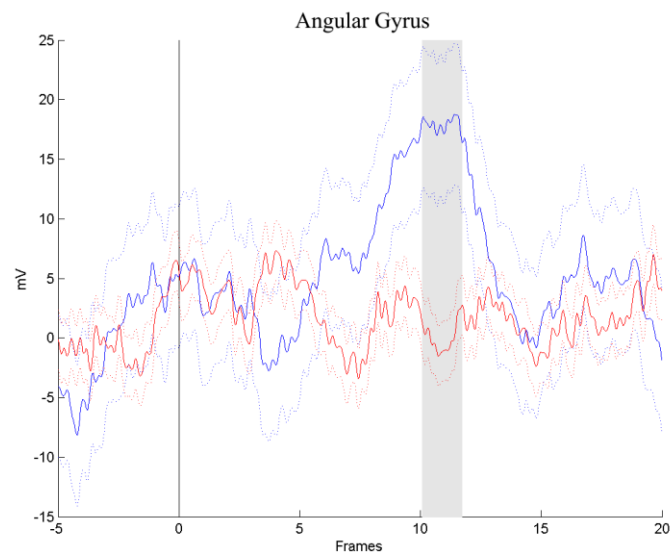
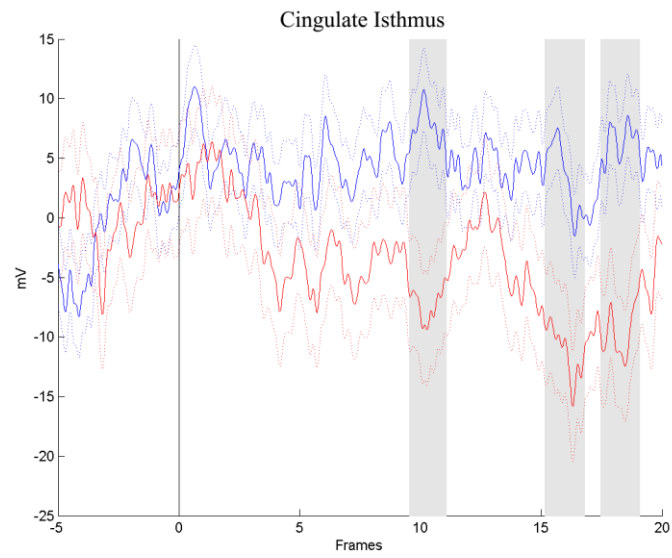
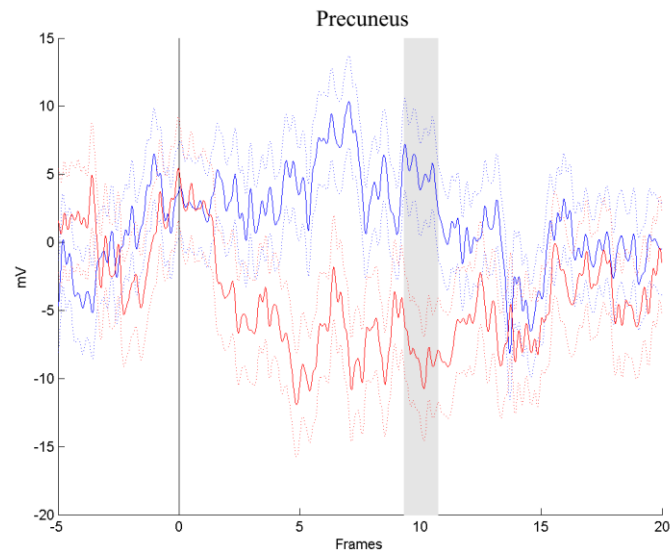


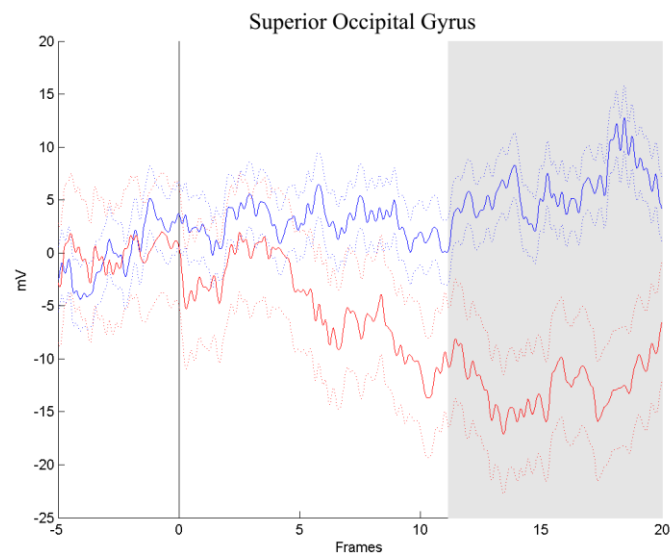
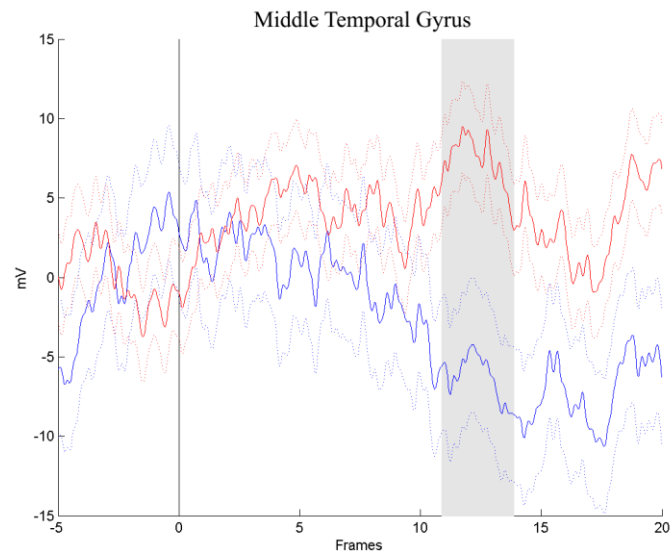
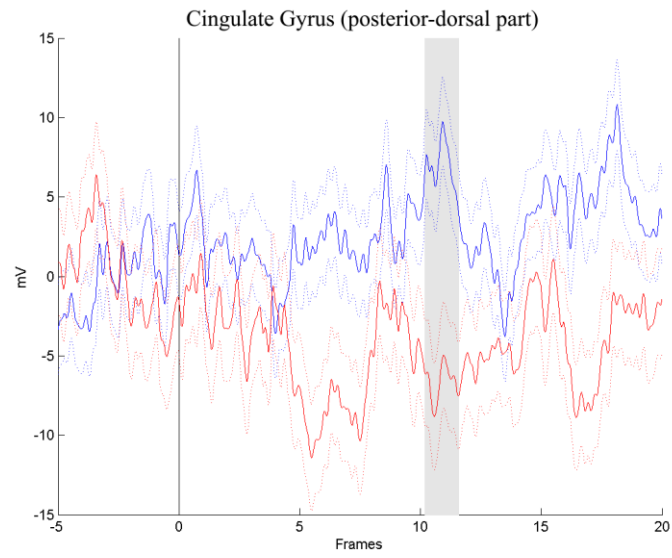
5.1.2 Speech with Seen Lips vs. Speech without Seen Lips

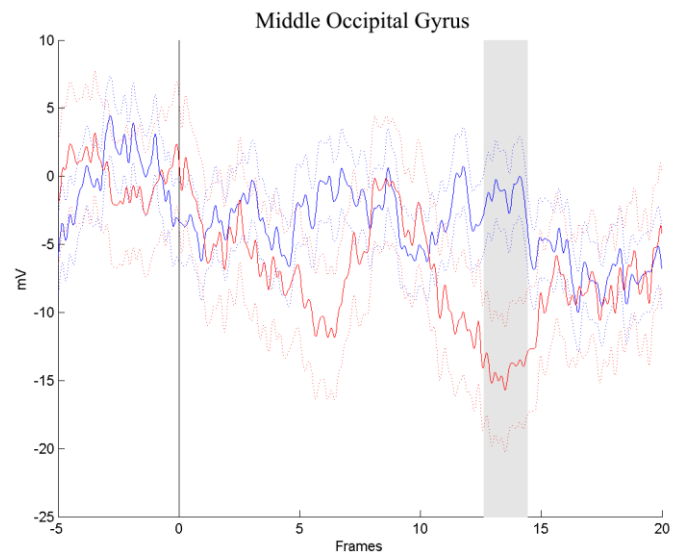
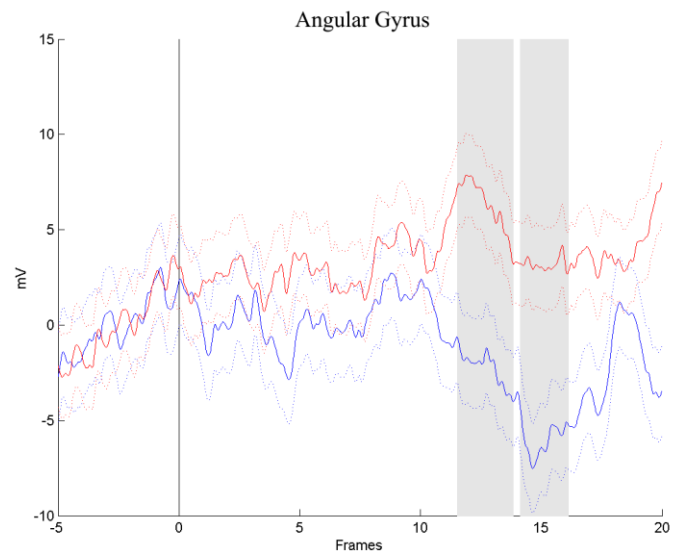
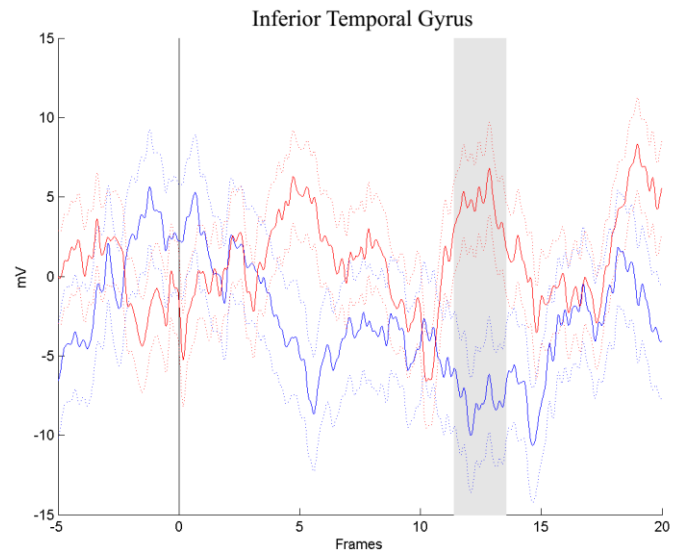
red: speech *with* seen lips; blue: speech *without* seen lips

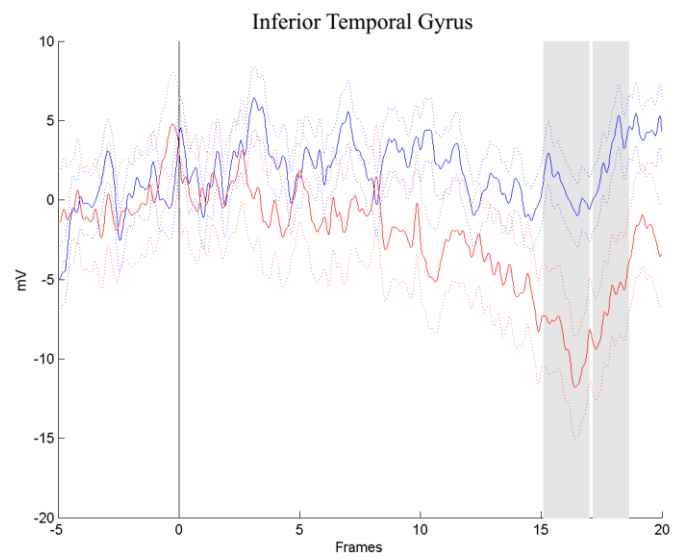
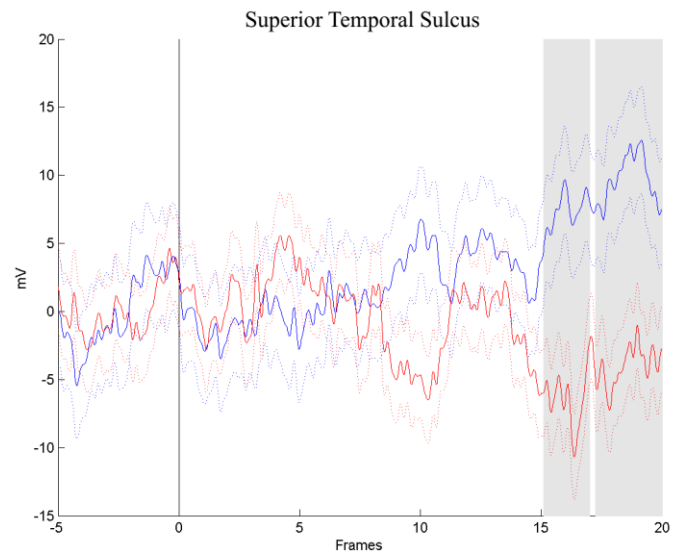
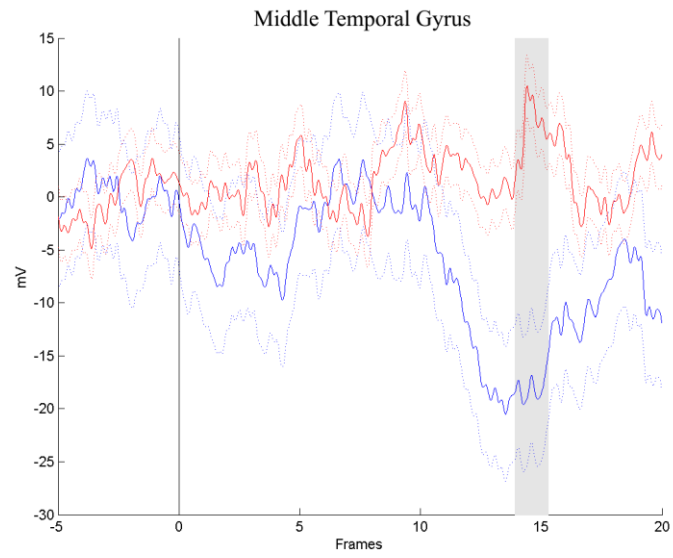


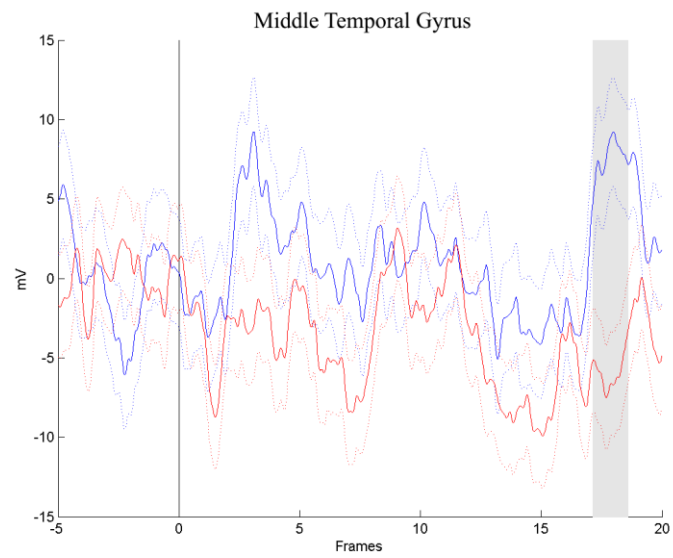
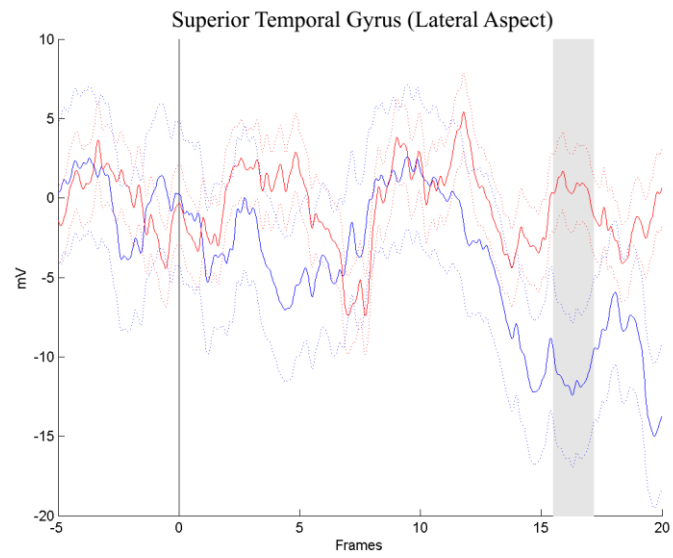
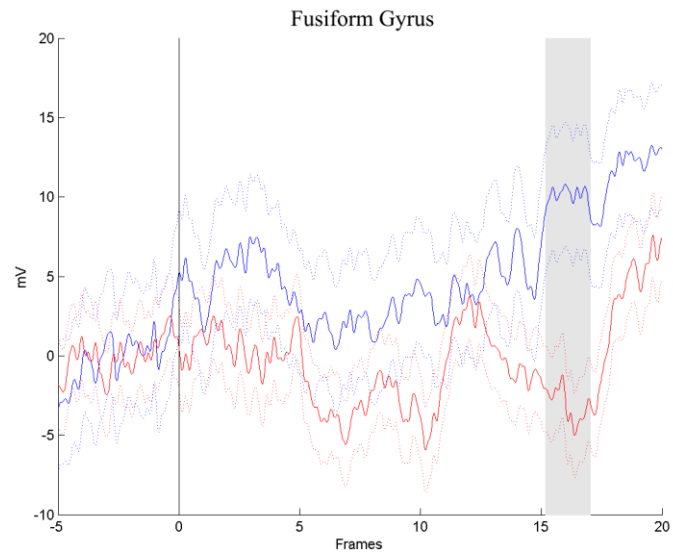


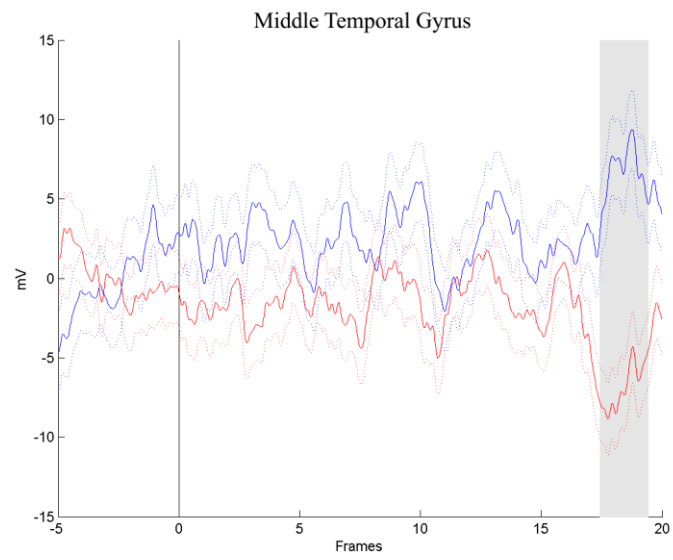












5.2 Abbreviations

5.2.1 Brain Regions

- BA: Brodmann Area
- dPM: dorsal Premotor
- ECFS: Extreme Capsule Fiber System
- FG: Fusiform Gyrus
- IPC: Inferior Parietal Cortex
- ITG: Inferior Temporal Gyrus
- MTG: Middle Temporal Gyrus
- PAC: Primary Auditory Cortex
- pIF: posterior Inferior Frontal²
- pITL: posterior Inferior Temporal Lobe
- PMC: Premotor Cortex
- PT: Planum Temporale (of the STG)
- SMA: Supplementary Motor Area
- SPL: Superior Parietal Lobe
- Spt: Sylvian-parietal-temporal
- STG: Superior Temporal Gyrus
- STS: Superior Temporal Sulcus

5.2.2 Imaging Techniques

- ECoG: Electrocorticography
- EEG: Electroencephalography
- fMRI: functional Magnetic Resonance Imaging
- MEG: Magnetoencephalography
- PET: Positron Emission Tomography

² In general, 'p' in front of an abbreviation of a brain region stands for 'posterior' and 'a' stands for 'anterior'.

5.2.3 Miscellaneous

- CI: Cochlear Implant
- dB: decibel(s)
- DFT: Discrete Fourier Transform
- mV: microvolt(s)

5.3 Algorithms

This section displays the most important MATLAB algorithms used for the data analyses in the experiment.

5.3.1 Statistical Test

This function performs a paired sample t-test between two groups of sounds given an α -value and a threshold for a period of significance. It then returns all periods of significance (start and end times).

```

1  function [sig_diff_starts , sig_diff_ends] = SignificantDiff(
2      pthresh , runthresh , envsound_clips , speech_clips)
3
4  sound_start = ceil(5*2000/24);
5
6  for s = 1:size(envsound_clips)
7      [~, pval(s)] = ttest2(envsound_clips(s, :), speech_clips(s, :));
8  end
9
10 hits = [0 pval<pthresh 0];
11 ons = (2+find(diff(hits) == 1));
12 offs = (2+find(diff(hits) == -1));
13 runsizes = offs - ons;
14 sig_diff_starts = [];
15 sig_diff_ends = [];
16 a = 1;
17
18 for i = 1:length(runsizes)
19     if (runsizes(i) >= runthresh) && (ons(i) >= sound_start)
20         sig_diff_starts(a) = ons(i);
21         % conversion to movie frames
22         sig_diff_starts(a) = sig_diff_starts(a)/2000*24-5;
23
24         sig_diff_ends(a) = offs(i);
25         % conversion to movie frames
26         sig_diff_ends(a) = sig_diff_ends(a)/2000*24-5;
27
28         a = a+1;
29     end
30 end

```

5.3.2 Loudness and Spectral Frequency Controls

The following code performs the controls for loudness and spectral frequency. In case of loudness, only those samples are kept whose peak loudness is below a certain threshold (in this example algorithm 80dB). To control for spectral frequency, the Discrete Fourier Transform (DFT) of every sample is computed and the samples that are most similar in spectral frequency, i.e. the samples with the smallest area between curves of the DFTs, are chosen. The responses to these samples are finally compared and plotted (see section 5.3.4).

```

1  function [envsound_times_quiet , speech_times_closestSF] =
      LoudnessSpecFreqControls(envsound_times , speech_times)
2
3  sound_length = 25; % frame number of observed window
4  sr = 48000;
5  sound_length_audio_samples = (sr*sound_length)/24;
6
7
8  % LOUDNESS CONTROL
9
10 envsound_clips_loudness = [];
11 envsound_clips_NewLoudness = [];
12 envsound_times_quiet = [];
13 a = 1;
14
15 for i = 1:length(envsound_times)
16     [envsound , sr] = wavread('Home_Alone_2_PG.wav' , [ ceil((48000*
          envsound_times(i))/24) ceil((48000*envsound_times(i))/24)+
          ceil(sound_length_audio_samples) -1], 'native');
17     envsound2 = max([abs(envsound(:,1)); abs(envsound(:,2))]);
18     envsound2 = double(envsound2);
19     envsound_clips_loudness(i) = mag2db(envsound2);
20
21     if(envsound_clips_loudness(i) <= 80)
22         envsound_times_quiet(a) = envsound_times(i);
23         envsound_clips_NewLoudness(a) = envsound_clips_loudness(i);
24         a = a+1;
25     end
26 end
27
28 speech_clips_loudness = [];
29 speech_clips_NewLoudness = [];
30 speech_times_quiet = [];
31 a = 1;
32
33 for i = 1:length(speech_times)
34     [speech , sr] = wavread('Home_Alone_2_PG.wav' , [ ceil((48000*
          speech_times(i))/24) ceil((48000*speech_times(i))/24)+ceil(
          sound_length_audio_samples) -1], 'native');
35     speech2 = max([abs(speech(:,1)); abs(speech(:,2))]);
36     speech2 = double(speech2);
37     speech_clips_loudness(i) = mag2db(speech2);
38
39     if(speech_clips_loudness(i) <= 80)
40         speech_times_quiet(a) = speech_times(i);
41         speech_clips_NewLoudness(a) = speech_clips_loudness(i);
42         a = a+1;
43     end
44 end
45

```

```

46
47 % SPECTRAL FREQUENCY CONTROL
48
49 envsound_clips_fft = nan(sound_length_audio_samples/2+1, length(
    envsound_times_quiet));
50
51 for i = 1:length(envsound_times_quiet)
52     [envsound, sr] = wavread('Home_Alone_2_PG.wav', [ceil((48000*
        envsound_times_quiet(i))/24) ceil((48000*
        envsound_times_quiet(i))/24)+ceil(
        sound_length_audio_samples)-1]);
53     envsound = abs(fft(envsound));
54     envsound = envsound(1:sound_length_audio_samples/2+1)/(
        sound_length_audio_samples/2+1);
55     envsound_clips_fft(:, i) = envsound;
56 end
57
58 speech_clips_fft = nan(sound_length_audio_samples/2+1, length(
    speech_times_quiet));
59
60 for i = 1:length(speech_times_quiet)
61     [speech, sr] = wavread('Home_Alone_2_PG.wav', [ceil((48000*
        speech_times_quiet(i))/24) ceil((48000*speech_times_quiet(i)
        ))/24)+ceil(sound_length_audio_samples)-1]);
62     speech = abs(fft(speech));
63     speech = speech(1:sound_length_audio_samples/2+1)/(
        sound_length_audio_samples/2+1);
64     speech_clips_fft(:, i) = speech;
65 end
66
67 closest_speechClip = [];
68 for i = 1:length(envsound_times_quiet)
69     clear areaBetweenCurves
70     for j = 1:length(speech_times_quiet)
71         areaBetweenCurves(j) = sum(abs( envsound_clips_fft(:, i)-
            speech_clips_fft(:, j) ));
72     end
73     mod_ABC = areaBetweenCurves;
74     mod_ABC(closest_speechClip) = inf;
75     [~, closest_speechClip(i)] = min(mod_ABC); %next
        closest_speechClip is the one with smallest
        areaBetweenCurves
76 end
77
78 closest_speech_clips_fft = [];
79 for i=1:length(closest_speechClip)
80     closest_speech_clips_fft(:, i) = speech_clips_fft(:,
        closest_speechClip(i));
81 end
82
83 speech_times_closestSF = [];
84 for i=1:length(closest_speechClip)
85     speech_times_closestSF(:, i) = speech_times_quiet(:,
        closest_speechClip(i));
86 end
87
88 % PLOT AVERAGE RESPONSES
89 func_Plot_MeanResponse_ALL(envsound_times_quiet ,
    speech_times_closestSF);

```


5.3.3 Threshold for Period of Significance

The threshold for a period of significance is computed using the subsequent code. We begin with a rather low threshold of 50 sampling points. For each electrode, half of the samples are then randomly assigned to one category and the other half to a second category. The responses to these samples are then compared. This process is performed 100 times. If a significant difference period is detected more than 5 times (out of the 100 times), then the threshold is increased by 10 sampling points. This procedure is repeated until a significant difference is found in less than 5 runs.³

```

1  load('m00055.Home Alone 2 (PG)_2_frame_data.mat')
2  b = [patient_clip_frame_samples{1}{:}, patient_clip_frame_samples
      {2}{:}];
3
4  sound_length = 25; % frames
5  sound_length_samples = (2000*sound_length)/24;
6
7  runThresholds = ones(144,1)*50;
8
9  all_times = horzcat(envsound_times_quiet, speech_times_closestSF);
10
11 for chan_num=1:144
12     % Exclusion of non-functioning electrodes
13     if (chan_num~=1 && chan_num~=10 && chan_num~=12 && chan_num~=13
        && chan_num~=15 && chan_num~=17 && chan_num~=18 &&
        chan_num~=19 && chan_num~=27 && chan_num~=28 && chan_num
        ~=29 && chan_num~=30 && chan_num~=32 && chan_num~=44)
14
15         chan = load ([ 'ch', num2str(chan_num), '.mat' ]);
16         sigDiffCounter = 100;
17
18         while(sigDiffCounter > 5)
19
20             sigDiffCounter = 0;
21             runThresholds(chan_num) = runThresholds(chan_num) + 10
22
23             for run=1:100
24
25                 % from all samples (all_times) randomly select 42 for
                % blue_times
26                 blue_times = randsample(all_times,42,false);
27                 % randomly select 42 for red_times
28                 red_times = randsample(all_times,42,false);
29
30                 % BLUE RESPONSES
31                 blue_clips = nan(ceil(sound_length_samples), length(
                blue_times));
32
33                 for i = 1:length(blue_times)
34                     blue_clips(:, i) = chan.ch(b(blue_times(i)):b(
                blue_times(i))+ceil(sound_length_samples)-1);
35                     blue_ranges(i) = max(blue_clips(:, i)) - min(blue_clips
                (:, i));
36                 end
37

```

³ The mean threshold where this was the case for all electrodes was taken as the threshold for a period of significance. This last step was done manually after the execution of the algorithm and is, thus, not performed in this algorithm.

```

38     for i = 1:length(blue_times)
39         if blue_ranges(i) > median(blue_ranges) + 4*std(
40             blue_ranges)
41             blue_clips(:, i) = nan;
42         end
43     end
44     blue_clips(isnan(blue_clips)) = nan;
45     blue_clips = reshape(blue_clips, ceil(
46         sound_length_samples), []);
47     mean_blue_clips = nanmean(blue_clips, 2);
48     % RED RESPONSES
49     red_clips = nan(ceil(sound_length_samples), length(
50         red_times));
51     for i = 1:length(red_times)
52         red_clips(:, i) = chan.ch(b(red_times(i)):b(red_times(i)
53             )+ceil(sound_length_samples)-1);
54         red_ranges(i) = max(red_clips(:, i)) - min(red_clips(:,
55             i));
56     end
57     for i = 1:length(red_times)
58         if red_ranges(i) > median(red_ranges) + 4*std(
59             red_ranges)
60             red_clips(:, i) = nan;
61         end
62     end
63     red_clips(isnan(red_clips)) = nan;
64     red_clips = reshape(red_clips, ceil(sound_length_samples)
65         , []);
66     mean_red_clips = nanmean(red_clips, 2);
67     % BASELINE CORRECTION
68     mean_baseline_red = nanmean(mean_red_clips(1 : ceil(
69         5*2000/24)));
70     mean_baseline_blue = nanmean(mean_blue_clips(1 : ceil(
71         5*2000/24)));
72     red_clips_tr = red_clips';
73     blue_clips_tr = blue_clips';
74     red_clips_tr = red_clips_tr - mean_baseline_red; %
75         subtract mean response in baseline period from all
76         samples
77     blue_clips_tr = blue_clips_tr - mean_baseline_blue;
78     mean_blue_clips = nanmean(blue_clips_tr);
79     mean_red_clips = nanmean(red_clips_tr);
80     % STATISTICAL TEST
81     [sig_diff_starts, sig_diff_ends] = SignificantDiff(0.05,
82         runThresholds(chan_num), blue_clips_tr', red_clips_tr');
83     ;
84     if (~isempty(sig_diff_starts))
85         sigDiffCounter = sigDiffCounter + 1
86     end
87     if (sigDiffCounter > 5)
88         break;
89     end
90 end

```

```

87     end
88     end
89 end

```

5.3.4 Comparison of Electrophysiological Responses

In this algorithm, which forms the core of the analyses, the electrode plots as seen in section 5.1 are created, as well as an array of the differentiation times of all electrodes that showed a significant difference between the two groups. In the process, abnormal responses are excluded and a baseline correction is performed. After sound onset, a "window" of 20 movie frames is observed. During this temporal window, the algorithm in 5.3.1 is used to test for statistically significant differences between the responses which are sustained over a period of at least 114 sampling points (the threshold for the period of significance was determined by the algorithm in 5.3.3).

```

1  function func_Plot_MeanResponse_ALL(blue_times, red_times)
2
3  load ('m00055_Home Alone 2 (PG)_2_frame_data.mat')
4  b = [patient_clip_frame_samples{1}{:},
      patient_clip_frame_samples{2}{:}];
5
6  sound_length = 25; %frames
7  sound_length_samples = (2000*sound_length)/24;
8  signDiffStarts = nan(144,5);
9  preview_length = 5;
10
11 for chan_num=1:144
12     % Exclusion of non-functioning electrodes
13     if (chan_num~=1 && chan_num~=10 && chan_num~=12 && chan_num
        ~=13 && chan_num~=15 && chan_num~=17 && chan_num~=18 &&
        chan_num~=19 && chan_num~=27 && chan_num~=28 && chan_num
        ~=29 && chan_num~=30 && chan_num~=32 && chan_num~=44)
14
15         chan = load (['ch', num2str(chan_num), '.mat']);
16
17         % BLUE RESPONSES
18         blue_clips = nan(ceil(sound_length_samples), length(
            blue_times));
19
20         for i = 1:length(blue_times)
21             blue_clips(:, i) = chan.ch(b(blue_times(i)):b(blue_times(i)
                )+ceil(sound_length_samples)-1);
22             blue_ranges(i) = max(blue_clips(:, i)) - min(blue_clips(:,
                i));
23         end
24
25         for i = 1:length(blue_times)
26             if blue_ranges(i) > median(blue_ranges) + 4*std(blue_ranges
                )
27                 blue_clips(:, i) = nan;
28             end
29         end
30
31         blue_clips(isnan(blue_clips)) = nan;
32         blue_clips = reshape(blue_clips, ceil(sound_length_samples),
            []);
33         mean_blue_clips = nanmean(blue_clips, 2);

```

```

34
35 % RED CLIPS
36 red_clips = nan(ceil(sound_length_samples), length(red_times)
    );
37
38 for i = 1:length(red_times)
39     red_clips(:, i) = chan.ch(b(red_times(i)):b(red_times(i))+
        ceil(sound_length_samples)-1);
40     red_ranges(i) = max(red_clips(:, i)) - min(red_clips(:, i))
        ;
41 end
42
43 for i = 1:length(red_times)
44     if red_ranges(i) > median(red_ranges) + 4*std(red_ranges)
45         red_clips(:, i) = nan;
46     end
47 end
48
49 red_clips(isnan(red_clips)) = nan;
50 red_clips = reshape(red_clips, ceil(sound_length_samples),
    []);
51 mean_red_clips = nanmean(red_clips, 2);
52
53 % BASELINE CORRECTION
54 mean_baseline_red = nanmean(mean_red_clips(1 : ceil
    (5*2000/24)));
55 mean_baseline_blue = nanmean(mean_blue_clips(1 : ceil
    (5*2000/24)));
56
57 red_clips_tr = red_clips';
58 blue_clips_tr = blue_clips';
59 red_clips_tr = red_clips_tr - mean_baseline_red; %subtract
    mean response in baseline period from all samples
60 blue_clips_tr = blue_clips_tr - mean_baseline_blue;
61 mean_blue_clips = nanmean(blue_clips_tr);
62 mean_red_clips = nanmean(red_clips_tr);
63
64 % PLOT
65 chan_plot = figure;
66 hold on;
67 ylabel('mV')
68 xlabel('Frames')
69 title(['Channel ', num2str(chan_num)])
70
71 plot((0:sound_length_samples)/2000*24-preview_length,
    mean_blue_clips, 'b-');
72 plot((0:sound_length_samples)/2000*24-preview_length,
    mean_blue_clips+std(mean_blue_clips), 'b:');
73 plot((0:sound_length_samples)/2000*24-preview_length,
    mean_blue_clips-std(mean_blue_clips), 'b:');
74
75 plot((0:sound_length_samples)/2000*24-preview_length,
    mean_red_clips, 'r-');
76 plot((0:sound_length_samples)/2000*24-preview_length,
    mean_red_clips+std(mean_red_clips), 'r:');
77 plot((0:sound_length_samples)/2000*24-preview_length,
    mean_red_clips-std(mean_red_clips), 'r:');
78
79 % STATISTICAL TEST
80 [sig_diff_starts, sig_diff_ends] = SignificantDiff(0.05, 114,
    blue_clips_tr', red_clips_tr') % computes periods of
    significant difference

```

```

81
82     yL = get(gca, 'YLim');
83     line ([0 0], yL, 'Color', 'k');
84     YLimits = ylim;
85
86     % Creation of grey rectangles showing the periods of
87     % significance
88     for i=1:length(sig_diff_starts)
89         rectangle ('Position', [sig_diff_starts(i), YLimits(1),
90             sig_diff_ends(i)-sig_diff_starts(i), abs(YLimits(1))+
91             YLimits(2)], 'FaceColor', [0.9,0.9,0.9], 'EdgeColor',
92             [0.9,0.9,0.9]);
93     end
94
95     % Mean response
96     plot ((0:sound_length_samples)/2000*24-preview_length,
97         mean_blue_clips, 'b-');
98     % Standard deviations
99     plot ((0:sound_length_samples)/2000*24-preview_length,
100         mean_blue_clips+std(mean_blue_clips), 'b:');
101     plot ((0:sound_length_samples)/2000*24-preview_length,
102         mean_blue_clips-std(mean_blue_clips), 'b:');
103
104     % Mean response
105     plot ((0:sound_length_samples)/2000*24-preview_length,
106         mean_red_clips, 'r-');
107     % Standard deviations
108     plot ((0:sound_length_samples)/2000*24-preview_length,
109         mean_red_clips+std(mean_red_clips), 'r:');
110     plot ((0:sound_length_samples)/2000*24-preview_length,
111         mean_red_clips-std(mean_red_clips), 'r:');
112
113     % Save plot as png
114     saveas(chan_plot, ['ch', num2str(chan_num), '.png'], 'png');
115     close all
116 end
117 end

```

BIBLIOGRAPHY

- Albert, M. L., Bachman, D. L., Morgan, A., and Helm-Estabrooks, N. (1988). Pharmacotherapy for aphasia. *Neurology*, 38(6):877–877.
- Altieri, N. (2010). *Toward a Unified Theory of Audiovisual Integration in Speech Perception*. Universal-Publishers.
- Altmann, G. and Gaskell, M. G. (2007). *The Oxford Handbook of Psycholinguistics*. Oxford University Press.
- Amunts, K., Schleicher, A., Bürgel, U., Mohlberg, H., Uylings, H., and Zilles, K. (1999). Broca’s region revisited: cytoarchitecture and intersubject variability. *Journal of Comparative Neurology*, 412(2):319–341.
- Asano, E., Juhász, C., Shah, A., Muzik, O., Chugani, D. C., Shah, J., Sood, S., and Chugani, H. T. (2005). Origin and propagation of epileptic spasms delineated on electrocorticography. *Epilepsia*, 46(7):1086–1097.
- Awad, I. A., Rosenfeld, J., Ahl, J., Hahn, J. F., and Lüders, H. (1991). Intractable epilepsy and structural lesions of the brain: mapping, resection strategies, and seizure outcome. *Epilepsia*, 32(2):179–186.
- Awh, E., Jonides, J., Smith, E. E., Schumacher, E. H., Koeppel, R. A., and Katz, S. (1996). Dissociation of storage and rehearsal in verbal working memory: Evidence from positron emission tomography. *Psychological Science*, 7:25–31.
- Baker, E., Blumstein, S. E., and Goodglass, H. (1981). Interaction between phonological and semantic factors in auditory comprehension. *Neuropsychologia*, 19(1):1–15.
- Belin, P., Fecteau, S., and Bédard, C. (2004). Thinking the voice: neural correlates of voice perception. *Trends in cognitive sciences*, 8(3):129–135.
- Belin, P. and Zatorre, R. J. (2003). Adaptation to speaker’s voice in right anterior temporal lobe. *Neuroreport*, 14(16):2105–2109.
- Belin, P., Zatorre, R. J., Lafaille, P., Ahad, P., and Pike, B. (2000). Voice-selective areas in human auditory cortex. *Nature*, 403(6767):309–312.
- Belin, P., Zilbovicius, M., Crozier, S., Thivard, L., Fontaine, A., Masure, M.-C., and Samson, Y. (1998). Lateralization of speech and auditory temporal processing. *Journal of cognitive neuroscience*, 10(4):536–540.
- Binder, J., Frost, J., Hammeke, T., Rao, S., and Cox, R. (1996). Function of the left planum temporale in auditory and linguistic processing. *Brain*, 119(4):1239–1247.

- Binder, J. R., Desai, R. H., Graves, W. W., and Conant, L. L. (2009). Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies. *Cerebral Cortex*, 19(12):2767–2796.
- Binder, J. R., Frost, J. A., Hammeke, T. A., Cox, R. W., Rao, S. M., and Prieto, T. (1997). Human brain language areas identified by functional magnetic resonance imaging. *The Journal of Neuroscience*, 17(1):353–362.
- Binder, J. R., Medler, D. A., Westbury, C. F., Liebenthal, E., and Buchanan, L. (2006). Tuning of the human left fusiform gyrus to sublexical orthographic structure. *Neuroimage*, 33(2):739–748.
- Binder, J. R., Rao, S. M., Hammeke, T., Yetkin, F., Jesmanowicz, A., Bandettini, P., Wong, E., Estkowski, L., Goldstein, M., Haughton, V., et al. (1994). Functional magnetic resonance imaging of human auditory cortex. *Annals of neurology*, 35(6):662–672.
- Bledowski, C., Prvulovic, D., Goebel, R., Zanella, F. E., and Linden, D. E. (2004). Attentional systems in target and distractor processing: a combined ERP and fMRI study. *Neuroimage*, 22(2):530–540.
- Boatman, D., Lesser, R. P., and Gordon, B. (1995). Auditory speech processing in the left temporal lobe: an electrical interference study. *Brain and language*, 51(2):269–290.
- Bornkessel, I. and Schlesewsky, M. (2006). The extended argument dependency model: a neurocognitive approach to sentence comprehension across languages. *Psychological review*, 113(4):787.
- Bornkessel, I., Zysset, S., Friederici, A. D., von Cramon, D. Y., and Schlesewsky, M. (2005). Who did what to whom? The neural basis of argument hierarchies during language comprehension. *Neuroimage*, 26(1):221–233.
- Broca, P. (1861). Remarques sur le siège de la faculté du langage articulé, suivies d'une observation daphémie (perte de la parole). *Bulletin de la Société Anatomique*, 6:330–357.
- Buchman, A. S., Garron, D. C., Trost-Cardamone, J. E., Wichter, M. D., and Schwartz, M. (1986). Word deafness: one hundred years later. *Journal of Neurology, Neurosurgery & Psychiatry*, 49(5):489–499.
- Buckner, R. L., Koutstaal, W., Schacter, D. L., and Rosen, B. R. (2000). Functional MRI evidence for a role of frontal and inferior temporal cortex in amodal components of priming. *Brain*, 123(3):620–640.
- Burton, M. W., Small, S. L., and Blumstein, S. E. (2000). The role of segmentation in phonological processing: an fMRI investigation. *Journal of Cognitive Neuroscience*, 12(4):679–690.
- Callan, D. E., Jones, J. A., Munhall, K., Kroos, C., Callan, A. M., and Vatikiotis-Bateson, E. (2004). Multisensory integration sites identified by perception of spatial wavelet filtered visual speech gesture information. *Journal of Cognitive Neuroscience*, 16(5):805–816.

- Calvert, G. and Lewis, J. (2003). Hemodynamic studies of audio-visual interactions.
- Calvert, G. A., Bullmore, E. T., Brammer, M. J., Campbell, R., Williams, S. C., McGuire, P. K., Woodruff, P. W., Iversen, S. D., and David, A. S. (1997). Activation of auditory cortex during silent lipreading. *Science*, 276(5312):593–596.
- Calvert, G. A. and Campbell, R. (2003). Reading speech from still and moving faces: the neural substrates of visible speech. *Journal of Cognitive Neuroscience*, 15(1):57–70.
- Calvert, G. A., Campbell, R., and Brammer, M. J. (2000). Evidence from functional magnetic resonance imaging of crossmodal binding in the human heteromodal cortex. *Current Biology*, 10(11):649–657.
- Campbell, R. (2008). The processing of audio-visual speech: empirical and neural bases. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 363(1493):1001–1010.
- Cao, F., Bitan, T., and Booth, J. R. (2008). Effective brain connectivity in children with reading difficulties during phonological processing. *Brain and language*, 107(2):91–101.
- Capek, C. M., Bavelier, D., Corina, D., Newman, A. J., Jezzard, P., and Neville, H. J. (2004). The cortical organization of audio-visual sentence comprehension: an fMRI study at 4 Tesla. *Cognitive brain research*, 20(2):111–119.
- Caplan, D., Alpert, N., and Waters, G. (1998). Effects of syntactic structure and propositional number on patterns of regional cerebral blood flow. *Journal of Cognitive Neuroscience*, 10(4):541–552.
- Caplan, D., Alpert, N., Waters, G., and Olivieri, A. (2000). Activation of Broca’s area by syntactic processing under conditions of concurrent articulation. *Human brain mapping*, 9(2):65–71.
- Caramazza, A. and Zurif, E. B. (1976). Dissociation of algorithmic and heuristic processes in language comprehension: Evidence from aphasia. *Brain and language*, 3(4):572–582.
- Catani, M., Jones, D. K., et al. (2005). Perisylvian language networks of the human brain. *Annals of neurology*, 57(1):8–16.
- Chee, M. W., O’Craven, K. M., Bergida, R., Rosen, B. R., and Savoy, R. L. (1999). Auditory and visual word processing studied with fMRI. *Human Brain Mapping*, 7(1):15–28.
- Creutzfeldt, O., Ojemann, G., and Lettich, E. (1989). Neuronal activity in the human lateral temporal lobe. *Experimental Brain Research*, 77(3):451–475.
- Crone, N. E., Boatman, D., Gordon, B., and Hao, L. (2001). Induced electrocorticographic gamma activity during auditory perception. *Clinical Neurophysiology*, 112(4):565–582.

- Damasio, A. R. (1992). Aphasia. *New England Journal of Medicine*, 326(8):531–539.
- Damasio, H. (1998). Neuroanatomical correlates of the aphasias. *Acquired aphasia*, page 43.
- Dapretto, M. and Bookheimer, S. Y. (1999). Form and content: dissociating syntax and semantics in sentence comprehension. *Neuron*, 24(2):427–432.
- Dehaene, S., Le Clec'H, G., Poline, J.-B., Le Bihan, D., and Cohen, L. (2002). The visual word form area: a prelexical representation of visual words in the fusiform gyrus. *Neuroreport*, 13(3):321–325.
- Demonet, J.-F., Chollet, F., Ramsay, S., Cardebat, D., Nespoulous, J.-L., Wise, R., Rascol, A., and Frackowiak, R. (1992). The anatomy of phonological and semantic processing in normal subjects. *Brain*, 115(6):1753–1768.
- Demonet, J.-F., Price, C., Wise, R., and Frackowiak, R. (1994). Differential activation of right and left posterior sylvian regions by semantic and phonological tasks: a positron-emission tomography study in normal human subjects. *Neuroscience letters*, 182(1):25–28.
- Desimone, R. and Gross, C. G. (1979). Visual areas in the temporal cortex of the macaque. *Brain research*, 178(2):363–380.
- Devlin, J., Jamison, H., Gonnerman, L., and Matthews, P. (2006). The role of the posterior fusiform gyrus in reading. *Cognitive Neuroscience, Journal of*, 18(6):911–922.
- DeWitt, I. and Rauschecker, J. P. (2012). Phoneme and word recognition in the auditory ventral stream. *Proceedings of the National Academy of Sciences*, 109(8):E505–E514.
- Dhankhar, A., Wexler, B. E., Fulbright, R. K., Halwes, T., Blamire, A. M., and Shulman, R. G. (1997). Functional magnetic resonance imaging assessment of the human brain auditory cortex response to increasing word presentation rates. *Journal of Neurophysiology*, 77(1):476–483.
- Dietz, N. A., Jones, K. M., Gareau, L., Zeffiro, T. A., and Eden, G. F. (2005). Phonological decoding involves left posterior fusiform gyrus. *Human brain mapping*, 26(2):81–93.
- Dodd, B. (1977). The role of vision in the perception of speech. *Perception*, 6(1):31–40.
- Ferstl, E. C. and von Cramon, D. Y. (2001). The role of coherence and cohesion in text comprehension: an event-related fMRI study. *Cognitive Brain Research*, 11(3):325–340.
- Ferstl, E. C. and von Cramon, D. Y. (2002). What does the frontomedian cortex contribute to language processing: coherence or theory of mind? *Neuroimage*, 17(3):1599–1612.

- Fiebach, C. J., Schlesewsky, M., and Friederici, A. D. (2001). Syntactic working memory and the establishment of filler-gap dependencies: Insights from ERPs and fMRI. *Journal of psycholinguistic research*, 30(3):321–338.
- Fiebach, C. J., Schlesewsky, M., Lohmann, G., Von Cramon, D. Y., and Friederici, A. D. (2005). Revisiting the role of Broca’s area in sentence processing: syntactic integration versus syntactic working memory. *Human brain mapping*, 24(2):79–91.
- Fiebach, C. J., Vos, S. H., and Friederici, A. D. (2004). Neural correlates of syntactic ambiguity in sentence comprehension for low and high span readers. *Journal of Cognitive Neuroscience*, 16(9):1562–1575.
- Fiez, J., Raichle, M., Miezin, F., Petersen, S., Tallal, P., and Katz, W. (1995). PET studies of auditory and phonological processing: Effects of stimulus characteristics and task demands. *Journal of Cognitive Neuroscience*, 7(3):357–375.
- Fletcher, P. C., Happe, F., Frith, U., Baker, S. C., Dolan, R. J., Frackowiak, R. S., and Frith, C. D. (1995). Other minds in the brain: a functional imaging study of ”theory of mind” in story comprehension. *Cognition*, 57(2):109–128.
- Friederici, A. D. (2002). Towards a neural basis of auditory sentence processing. *Trends in cognitive sciences*, 6(2):78–84.
- Friederici, A. D. (2004). Processing local transitions versus long-distance syntactic hierarchies. *Trends in cognitive sciences*, 8(6):245–247.
- Friederici, A. D. (2009). Pathways to language: fiber tracts in the human brain. *Trends in cognitive sciences*, 13(4):175–181.
- Friederici, A. D. (2011). The brain basis of language processing: from structure to function. *Physiological reviews*, 91(4):1357–1392.
- Friederici, A. D. (2012). The cortical language circuit: from auditory perception to sentence comprehension. *Trends in cognitive sciences*, 16(5):262–268.
- Friederici, A. D., Bahlmann, J., Heim, S., Schubotz, R. I., and Anwander, A. (2006). The brain differentiates human and non-human grammars: functional localization and structural connectivity. *Proceedings of the National Academy of Sciences of the United States of America*, 103(7):2458–2463.
- Friederici, A. D., Meyer, M., and von Cramon, D. Y. (2000). Auditory language comprehension: an event-related fMRI study on the processing of syntactic and lexical information. *Brain and language*, 74(2):289–300.
- Gallagher, H. L., Happé, F., Brunswick, N., Fletcher, P. C., Frith, U., and Frith, C. D. (2000). Reading the mind in cartoons and stories: an fMRI study of ’theory of mind’ in verbal and nonverbal tasks. *Neuropsychologia*, 38(1):11–21.
- Geschwind, N. (1970). The organization of language and the brain. In *Science*. Citeseer.

- Giraud, A. L. and Price, C. (2001). The constraints functional neuroimaging places on classical models of auditory word processing. *Journal of cognitive neuroscience*, 13(6):754–765.
- Giraud, A.-L., Price, C. J., Graham, J. M., Truy, E., and Frackowiak, R. S. (2001). Cross-modal plasticity underpins language recovery after cochlear implantation. *Neuron*, 30(3):657–664.
- Giraud, A. L. and Truy, E. (2002). The contribution of visual areas to speech comprehension: a PET study in cochlear implants patients and normal-hearing subjects. *Neuropsychologia*, 40(9):1562–1569.
- Goodale, M. A. and Milner, A. D. (1992). Separate visual pathways for perception and action. *Trends in neurosciences*, 15(1):20–25.
- Grodzinsky, Y. (2000). The neurology of syntax: Language use without Broca’s area. *Behavioral and brain sciences*, 23(01):1–21.
- Gruber, T. and Müller, M. M. (2005). Oscillatory brain activity dissociates between associative stimulus content in a repetition priming task in the human EEG. *Cerebral Cortex*, 15(1):109–116.
- Hickok, G., Buchsbaum, B., Humphries, C., and Muftuler, T. (2003). Auditory–motor interaction revealed by fMRI: speech, music, and working memory in area Spt. *Journal of Cognitive Neuroscience*, 15(5):673–682.
- Hickok, G. and Poeppel, D. (2000). Towards a functional neuroanatomy of speech perception. *Trends in cognitive sciences*, 4(4):131–138.
- Hickok, G. and Poeppel, D. (2004). Dorsal and ventral streams: a framework for understanding aspects of the functional anatomy of language. *Cognition*, 92(1):67–99.
- Hickok, G. and Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews Neuroscience*, 8(5):393–402.
- Hikosaka, K., Iwai, E., Saito, H., and Tanaka, K. (1988). Polysensory properties of neurons in the anterior bank of the caudal superior temporal sulcus of the macaque monkey. *Journal of neurophysiology*, 60(5):1615–1637.
- Horwitz, B., Rumsey, J. M., and Donohue, B. C. (1998). Functional connectivity of the angular gyrus in normal reading and dyslexia. *Proceedings of the National Academy of Sciences*, 95(15):8939–8944.
- Jones, E. and Powell, T. (1970). An anatomical study of converging sensory pathways within the cerebral cortex of the monkey. *Brain*, 93(4):793–820.
- Kaan, E. and Swaab, T. Y. (2002). The brain circuitry of syntactic comprehension. *Trends in cognitive sciences*, 6(8):350–356.
- Kertesz, A., Sheppard, A., and MacKenzie, R. (1982). Localization in transcortical sensory aphasia. *Archives of Neurology*, 39(8):475.

- Kim, J.-J., Crespo-Facorro, B., Andreasen, N. C., O’Leary, D. S., Magnotta, V., and Nopoulos, P. (2003). Morphology of the lateral superior temporal gyrus in neuroleptic naïve patients with schizophrenia: relationship to symptoms. *Schizophrenia research*, 60(2):173–181.
- Kuruvilla, A. and Flink, R. (2003). Intraoperative electrocorticography in epilepsy surgery: useful or not? *Seizure*, 12(8):577–584.
- Lau, E. F., Phillips, C., and Poeppel, D. (2008). A cortical network for semantics: (de)constructing the N400. *Nature Reviews Neuroscience*, 9(12):920–933.
- Leaver, A. M. and Rauschecker, J. P. (2010). Cortical representation of natural complex sounds: effects of acoustic features and auditory object category. *The Journal of Neuroscience*, 30(22):7604–7612.
- Levelt, W. J. (1999). Models of word production. *Trends in cognitive sciences*, 3(6):223–232.
- Logothetis, N. K. (2003). The underpinnings of the BOLD functional magnetic resonance imaging signal. *The Journal of Neuroscience*, 23(10):3963–3971.
- MacDonald, J. and McGurk, H. (1978). Visual influences on speech perception processes. *Perception & Psychophysics*, 24(3):253–257.
- Mark, F., Barry, W., and Michael, A. (2001). Neuroscience: Exploring the Brain. *Maryland: Lippincott williams and wilkins*.
- Mazoyer, B. M., Tzourio, N., Frak, V., Syrota, A., Murayama, N., Levrier, O., Salamon, G., Dehaene, S., Cohen, L., and Mehler, J. (1993). The cortical representation of speech. *Journal of Cognitive Neuroscience*, 5(4):467–479.
- McCandliss, B. D., Cohen, L., and Dehaene, S. (2003). The visual word form area: expertise for reading in the fusiform gyrus. *Trends in cognitive sciences*, 7(7):293–299.
- McGlone, J. (1984). Speech comprehension after unilateral injection of sodium amytal. *Brain and Language*, 22(1):150–157.
- McGurk, H. and MacDonald, J. (1976). Hearing lips and seeing voices.
- Mechelli, A., Humphreys, G. W., Mayall, K., Olson, A., and Price, C. J. (2000). Differential effects of word length and visual contrast in the fusiform and lingual gyri during reading. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 267(1455):1909–1913.
- Mesgarani, N. and Chang, E. F. (2012). Selective cortical representation of attended speaker in multi-talker speech perception. *Nature*, 485(7397):233–236.
- Meyer, M., Alter, K., and Friederici, A. (2003). Functional MR imaging exposes differential brain responses to syntax and prosody during auditory sentence comprehension. *Journal of Neurolinguistics*, 16(4):277–300.
- Meyer, M., Friederici, A. D., and von Cramon, D. Y. (2000). Neurocognition of auditory sentence comprehension: event related fMRI reveals sensitivity to syntactic violations and task demands. *Cognitive Brain Research*, 9(1):19–33.

- Miller, K. J., denNijs, M., Shenoy, P., Miller, J. W., Rao, R. P., and Ojemann, J. G. (2007). Real-time functional brain mapping using electrocorticography. *Neuroimage*, 37(2):504–507.
- Mulert, C., Jäger, L., Schmitt, R., Bussfeld, P., Pogarell, O., Möller, H.-J., Juckel, G., and Hegerl, U. (2004). Integration of fMRI and simultaneous EEG: towards a comprehensive understanding of localization and time-course of brain activity in target detection. *Neuroimage*, 22(1):83–94.
- Müller, R.-A., Kleinhans, N., and Courchesne, E. (2003). Linguistic theory and neuroimaging evidence: An fMRI study of Broca’s area in lexical semantics. *Neuropsychologia*, 41(9):1199–1207.
- Müller, R.-A., Rothermel, R. D., Behen, M. E., Muzik, O., Mangner, T. J., and Chugani, H. T. (1997). Receptive and expressive language activations for sentences: a PET study. *Neuroreport*, 8(17):3767–3770.
- Newman, S. D., Ikuta, T., and Burns Jr, T. (2010). The effect of semantic relatedness on syntactic analysis: an fMRI study. *Brain and language*, 113(2):51–58.
- Nishitani, N. and Hari, R. (2002). Viewing lip forms: cortical dynamics. *Neuron*, 36(6):1211–1220.
- Obleser, J. and Kotz, S. A. (2010). Expectancy constraints in degraded speech modulate the language comprehension network. *Cerebral Cortex*, 20(3):633–640.
- Obleser, J., Wise, R. J., Dresner, M. A., and Scott, S. K. (2007a). Functional integration across brain regions improves speech perception under adverse listening conditions. *The Journal of Neuroscience*, 27(9):2283–2289.
- Obleser, J., Zimmermann, J., Van Meter, J., and Rauschecker, J. P. (2007b). Multiple stages of auditory speech perception reflected in event-related fMRI. *Cerebral Cortex*, 17(10):2251–2257.
- O’Grady, W., Dobrovolsky, M., and Katamba, F. (1997). *Contemporary linguistics: An introduction*. Longman Harlow, UK.
- Patterson, K., Nestor, P. J., and Rogers, T. T. (2007). Where do you know what you know? The representation of semantic knowledge in the human brain. *Nature Reviews Neuroscience*, 8(12):976–987.
- Paus, T., Perry, D. W., Zatorre, R. J., Worsley, K. J., and Evans, A. C. (1996). Modulation of cerebral blood flow in the human auditory cortex during speech: Role of motor-to-sensory discharges. *European Journal of Neuroscience*, 8(11):2236–2246.
- Pekkola, J., Ojanen, V., Autti, T., Jääskeläinen, I. P., Möttönen, R., Tarkiainen, A., and Sams, M. (2005). Primary auditory cortex activation by visual speech: an fMRI study at 3 T. *Neuroreport*, 16(2):125–128.
- Penfield, W. and Jasper, H. (1954). Epilepsy and the functional anatomy of the human brain.

- Perani, D., Paulesu, E., Galles, N. S., Dupoux, E., Dehaene, S., Bettinardi, V., Cappa, S. F., Fazio, F., and Mehler, J. (1998). The bilingual brain. proficiency and age of acquisition of the second language. *Brain*, 121(10):1841–1852.
- Petersen, S. E., Fox, P. T., Posner, M. I., Mintun, M., and Raichle, M. E. (1988). Positron emission tomographic studies of the cortical anatomy of single-word processing. *Nature*, 331(6157):585–589.
- Price, C., Wise, R., Warburton, E., Moore, C., Howard, D., Patterson, K., Frackowiak, R., and Friston, K. (1996). Hearing and saying: The functional neuro-anatomy of auditory word processing. *Brain*, 119(3):919–931.
- Price, C. J. (2000). The anatomy of language: contributions from functional neuroimaging. *Journal of anatomy*, 197(3):335–359.
- Rauschecker, J. P. (1998). Cortical processing of complex sounds. *Current opinion in neurobiology*, 8(4):516–521.
- Rauschecker, J. P. and Scott, S. K. (2009). Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. *Nature neuroscience*, 12(6):718–724.
- Robin, D. A., Tranel, D., and Damasio, H. (1990). Auditory perception of temporal and spectral events in patients with focal left and right cerebral lesions. *Brain and language*, 39(4):539–555.
- Rodd, J. M., Davis, M. H., and Johnsrude, I. S. (2005). The neural mechanisms of speech comprehension: fMRI studies of semantic ambiguity. *Cerebral cortex*, 15(8):1261–1269.
- Sams, M., Aulanko, R., Hämäläinen, M., Hari, R., Lounasmaa, O. V., Lu, S.-T., and Simola, J. (1991). Seeing speech: visual information from lip movements modifies activity in the human auditory cortex. *Neuroscience letters*, 127(1):141–145.
- Santi, A. and Grodzinsky, Y. (2010). fMRI adaptation dissociates syntactic complexity dimensions. *Neuroimage*, 51(4):1285–1293.
- Saur, D., Kreher, B. W., Schnell, S., Kümmerer, D., Kellmeyer, P., Vry, M.-S., Umarova, R., Musso, M., Glauche, V., Abel, S., et al. (2008). Ventral and dorsal pathways for language. *Proceedings of the National Academy of Sciences*, 105(46):18035–18040.
- Schlosser, M. J., Aoyagi, N., Fulbright, R. K., Gore, J. C., and McCarthy, G. (1998). Functional MRI studies of auditory comprehension. *Human brain mapping*, 6(1):1–13.
- Schuh, L. and Drury, I. (1997). Intraoperative electrocorticography and direct cortical electrical stimulation. In *Seminars in Anesthesia, Perioperative Medicine and Pain*, volume 16, pages 46–55. Elsevier.
- Scott, S. K., Blank, C. C., Rosen, S., and Wise, R. J. (2000). Identification of a pathway for intelligible speech in the left temporal lobe. *Brain*, 123(12):2400–2406.

- Scott, S. K. and Wise, R. J. (2004). The functional neuroanatomy of prelexical processing in speech perception. *Cognition*, 92(1):13–45.
- Sekiyama, K., Kanno, I., Miura, S., and Sugita, Y. (2003). Auditory-visual speech perception examined by fMRI and PET. *Neuroscience research*, 47(3):277–287.
- Seltzer, B. and Pandya, D. N. (1978). Afferent cortical connections and architectonics of the superior temporal sulcus and surrounding cortex in the rhesus monkey. *Brain research*, 149(1):1–24.
- Seth, A. K. (2010). A MATLAB toolbox for Granger causal connectivity analysis. *Journal of neuroscience methods*, 186(2):262–273.
- Shaywitz, B. A., Shaywitz, S. E., Pugh, K. R., Mencl, W. E., Fulbright, R. K., Skudlarski, P., Constable, R. T., Marchione, K. E., Fletcher, J. M., Lyon, G. R., et al. (2002). Disruption of posterior brain systems for reading in children with developmental dyslexia. *Biological psychiatry*, 52(2):101–110.
- Skipper, J. I., van Wassenhove, V., Nusbaum, H. C., and Small, S. L. (2007). Hearing lips and seeing voices: how cortical areas supporting speech production mediate audiovisual speech perception. *Cerebral Cortex*, 17(10):2387–2399.
- Spencer, K. M., Dien, J., and Donchin, E. (1999). A componential analysis of the ERP elicited by novel events using a dense electrode array. *Psychophysiology*, 36(03):409–414.
- Stephan, A. and Walter, S., editors (2013). *Handbuch Kognitionswissenschaft*. J.B. Metzler.
- Stromswold, K., Caplan, D., Alpert, N., and Rauch, S. (1996). Localization of syntactic comprehension by positron emission tomography. *Brain and language*, 52(3):452–473.
- Sugano, H., Shimizu, H., and Sunaga, S. (2007). Efficacy of intraoperative electrocorticography for assessing seizure outcomes in intractable epilepsy patients with temporal-lobe-mass lesions. *Seizure*, 16(2):120–127.
- Sumby, W. and Pollack, I. (1954). Visual contribution to speech intelligibility in noise. *Journal of the Acoustical Society of America*, 26:212–215.
- Tyler, L. K., Marslen-Wilson, W. D., Randall, B., Wright, P., Devereux, B. J., Zhuang, J., Papoutsis, M., and Stamatakis, E. A. (2011). Left inferior frontal cortex and syntax: function, structure and behaviour in patients with left hemisphere damage. *Brain*, 134(2):415–431.
- Vandenberghe, R., Nobre, A., and Price, C. (2002). The response of left temporal cortex to sentences. *Journal of Cognitive Neuroscience*, 14(4):550–560.
- Vitacco, D., Brandeis, D., Pascual-Marqui, R., and Martin, E. (2002). Correspondence of event-related potential tomography and functional magnetic resonance imaging during language processing. *Human brain mapping*, 17(1):4–12.

- Wada, J. and Rasmussen, T. (1960). Intracarotid injection of sodium amytal for the lateralization of cerebral speech dominance: experimental and clinical observations. *Journal of Neurosurgery*, 17(2):266–282.
- Ward, J. (2006). The student’s guide to cognitive neuroscience. *Hove and New York*.
- Warren, J., Scott, S., Price, C., and Griffiths, T. (2006). Human brain mechanisms for the early analysis of voices. *Neuroimage*, 31(3):1389–1397.
- Weiller, C., Musso, M., Rijntjes, M., and Saur, D. (2009). Please don’t underestimate the ventral pathway in language. *Trends in cognitive sciences*, 13(9):369–370.
- Wernicke, C. (1874). *Der aphasische Symptomenkomplex: eine psychologische Studie auf anatomischer Basis*. Springer.
- Westbury, C., Zatorre, R., and Evans, A. (1999). Quantifying variability in the planum temporale: A probability map. *Cerebral Cortex*, 9(4):392–405.
- Wilson, R. A. and Keil, F. C. (2001). *The MIT encyclopedia of the cognitive sciences*. MIT press.
- Wilson, S. M., Dronkers, N. F., Ogar, J. M., Jang, J., Growdon, M. E., Agosta, F., Henry, M. L., Miller, B. L., and Gorno-Tempini, M. L. (2010). Neural correlates of syntactic processing in the nonfluent variant of primary progressive aphasia. *The Journal of Neuroscience*, 30(50):16845–16854.
- Wise, R. J., Howard, D., Mummery, C., Fletcher, P., Leff, A., Büchel, C., and Scott, S. (2000). Noun imageability and the temporal lobes. *Neuropsychologia*, 38(7):985–994.
- Wise, R. J., Scott, S. K., Blank, S. C., Mummery, C. J., Murphy, K., and Warburton, E. A. (2001). Separate neural subsystems within ‘Wernicke’s area’. *Brain*, 124(1):83–95.
- Xu, J., Kemeny, S., Park, G., Frattali, C., and Braun, A. (2005). Language in context: emergent features of word, sentence, and narrative comprehension. *Neuroimage*, 25(3):1002–1015.
- Zaidel, E. (1985). Language in the right hemisphere.
- Zatorre, R. J. (1997). Cerebral correlates of human auditory processing. In *Acoustical signal processing in the central auditory system*, pages 453–468. Springer.
- Zatorre, R. J., Evans, A. C., Meyer, E., and Gjedde, A. (1992). Lateralization of phonetic and pitch discrimination in speech processing. *Science*, 256(5058):846–849.
- Zatorre, R. J., Meyer, E., Gjedde, A., and Evans, A. C. (1996). PET studies of phonetic processing of speech: review, replication, and reanalysis. *Cerebral cortex*, 6(1):21–30.
- Zekveld, A. A., Heslenfeld, D. J., Festen, J. M., and Schoonhoven, R. (2006). Top-down and bottom-up processes in speech comprehension. *Neuroimage*, 32(4):1826–1836.

Declaration of Authorship

I hereby certify that the work presented here is, to the best of my knowledge and belief, original and the result of my own investigations, except as acknowledged, and has not been submitted, either in part or whole, for a degree at this or any other university.

signature

city, date