Jussi Alho

The role of the speech motor system in speech perception

Faculty of Electronics, Communications and Automation

In partial fulfillment of the requirements for the degree of Master of Science in Technology, Espoo May 26, 2010.

Supervisor: Professor Mikko Sams

Instructor: Ph.D. Iiro Jääskeläinen

AALTO YLIOPISTO
TEKNILLINEN KORKEAKOULU

DIPLOMITYÖN
TIIVISTELMÄ

Tekijä: Jussi Alho

Otsikko: Puheen motorisen järjestelmän rooli puheen havaitsemisessa

Päivämäärä: 26.5.2010 Kieli: englanti Sivumäärä: 80

Elektroniikan, tietoliikenteen ja automaation tiedekunta

Professuuri: Kognitiivinen teknologia Koodi: S-114

Valvoja: Prof. Mikko Sams

Ohjaaja: FT Iiro Jääskeläinen

Tiivistelmäteksti:

Puheen havaitsemisella viitataan prosessiin, jonka puitteissa kielen äänteet kuullaan, tulkitaan ja ymmärretään. Vaikka tämä prosessi saattaa kuulostaa mitäänsanomattoman yksinkertaiselta, ovat ne neuraaliset mekanismit, jotka sen mahdollistavat, kaikkea muuta kuin yksinkertaisia. Keskeinen tutkijoiden päitä vaivaava kysymys puheen havaitsemisessa onkin, miten kuulija poimii olennaisen informaation puhesignaalista? Vastauksen etsintä on synnyttänyt toisistaan selvästi erillään olevia teoreettisia näkökulmia, joiden keskeisin erottava tekijä voidaan esittää kysymyksen muodossa: mikä on puheen motorisen järjestelmän rooli puheen havaitsemisessa? Vai onko sillä roolia laisinkaan? Toisin sanoen, onko niillä aivoalueilla ja –rakenteilla, jotka vastaavat puheen tuottamisesta, osuutta myös puheen havaitsemisessa?

Tässä työssä puheen motorisen järjestelmän roolia puheen havaitsemisessa tutkittiin käyttäen tutkimusmenetelmänä magnetoenkefalografiaa (MEG). Puheen havaitsemisen perustana olevia neuraalisia mekanismeja tutkittiin neljässä erilaisessa havaitsemistilanteessa, joissa jokaisessa käytettiin samoja kahden eri selkeystason puheärsykkeitä (kohinalla ja ilman), mutta joissa koehenkilöiltä edellytettiin erilaisia ärsykkeenjälkeisiä motorisia toimenpiteitä.

Yhteensä kymmeneltä koehenkilöltä saadut tulokset, joissa aivoaktivaation lähteitä mallinnettiin yhden dipolin analyysimenetelmällä, osoittavat että ”aktivaation painopiste” vasemmassa aivopuoliskossa siirtyi taaemmaksi kuuloaivokuorella havaitsemistilanteiden motoristen toimenpiteiden aktiivisuuden lisääntyessä. Tämä löydös puoltaa ajatusta siitä, että puheen motorisella järjestelmällä on rooli puheen havaitsemisessa.

Tämä työ, joka suoritettiin valmistelevana osana laajempaa projektia, luo pohjatyöllään ja tuloksillaan lupaavat lähtökohdat mitä erilaisimpiin puheen havaitsemisen neuraalista luonnetta valottaviin jatkoanalyyseihin.

Avainsanat: magnetoenkefalografia, puheen havaitseminen, herätekentät, puheen motorinen teoria, kuuloaivokuori

Author: Jussi Alho

Title: Role of the speech motor system in speech perception

Date: 26.5.2010 **Language**: english **Number of pages**: 80

Faculty of Electronics, Communications and Automation

Professorship: Cognitive technology **Code**: S-114

Supervisor: Prof. Mikko Sams

Instructor: Ph.D. Iiro Jääskeläinen

Abstract:

Speech perception refers to a process by which the sounds of language are heard, interpreted and understood. Although this process may seem like a trivial task, the neural mechanisms underlying it are anything but simple. Indeed, the key question puzzling the minds of speech perception researchers is how listeners extract the significant information from the acoustic speech signal? How is the mapping between properties of the acoustic signal and linguistic elements, such as phonemes, done? Distinct theoretical perspectives have been proposed to answer these questions. A crucial distinction among these perspectives can be put in the form of a question: does the speech motor system have a role in speech perception?

In this thesis, the role of the speech motor system in speech perception was studied using magnetoencephalography (MEG). The neural mechanisms underlying speech perception was investigated in four different perception conditions, each comprising the same speech stimuli (with two levels of ambiguity) but differing on the subsequent motor task.

The results, derived from ten subjects, show a clear shift of the equivalent current dipole (ECD), used in modeling the underlying neuronal sources, to a more posterior position in the left hemisphere with the more active subsequent-to-stimuli motor tasks. This suggests that the motor system does indeed have a role in speech perception.

The outcome of this thesis, which was conducted as a preliminary part of a larger project, serves as a promising basis for further study on speech perception, providing the necessary groundwork which allows more refined analyses to take place.

Keywords: magnetoencephalography, speech perception, event-related fields, motor theory of speech perception, auditory cortex

Foreword

This work was conducted in the Department of Biomedical Engineering and Computational Science (BECS) in the Aalto University School of Science and Technology. The supervisor of this thesis was Prof. Mikko Sams and the instructor was Ph.D. Iiro Jääskeläinen.

I would like to thank both my instructor Ph.D. Iiro Jääskeläinen and my supervisor Prof. Mikko Sams for the opportunity to conduct this thesis at BECS. Furthermore, I wish to express my gratitude for their inspiring expertise as well as for the trust they have bestowed on me. Especially the constructive comments and directions from my instructor were crucial in guiding me through the research and writing processes. Additional thanks go to M.Sc. Jaakko Kauramäki for his valuable guidance throughout the process, to Ph.D. Hannu Tiitinen for his instructions on MEG recordings, and to Ph.D. Jean-Luc Schwartz and Ph.D. Marc Sato for the collaboration and groundwork concerning the project within which this thesis was carried out, with a special mention to Ph.D. Marc Sato for providing me with essential background material on the project.

Finally, I would like to thank my parents for the support they have given me throughout my studies. I also wish to thank all the volunteers who agreed to subject themselves for science. Praise also to all my friends for helping me to relax and give my mind an occasional well deserved rest from the thesis work.

In Espoo, 26th May 2010

Contents

List of figures

List of tables

Abbreviations

A	auditory
a	anterior
AEF	auditory evoked field
AEP	auditory evoked potential
AV	audiovisual
BA	Broca's area
DRT	direct realist theory (of speech perception)
ECD	equivalent current dipole
EEG	electroencephalography
EMG	electromyography
EOG	electro-oculography
ERF	event-related field
ERP	event-related potential
fMRI	functional magnetic resonance imaging
HG	Heschl's gyrus
IFG	inferior frontal gyrus
ISI	interstimulus interval
M1	primary motor cortex
MEG	magnetoencephalography
MGN	medial geniculate nucleus
MT	motor theory (of speech perception)
MTG	middle temporal gyrus
N100m	negative peak in the ERF approximately 100 ms from the stimulus onset
p	posterior
P100m	positive peak in the ERf approximately 100 ms from the stimulus onset
P200m	positive peak in the ERf approximately 200 ms from the stimulus onset
PAC	primary auditory cortex
PET	positron emission tomography
PMC	premotor cortex
PT	planum temporale
rTMS	repetitive transcranial magnetic stimulation
SEM	standard error of the mean

SNR	signal-to-noise ratio
SQUID	superconducting quantum interference device
STG	superior temporal gyrus
STS	superior temporal sulcus
svPMC	left ventral premotor cortex
TMS	transcranial magnetic stimulation
V	visual
v	ventral

Chapter 1

Introduction

For most of us, perceiving speech is an effortless task and may thus seem like a trivial process. Is it though? Let's start by defining what exactly is meant by speech perception. One possible definition comes from Wikipedia ("Speech perception", 2010): "Speech perception is the process by which the sounds of language are heard, interpreted and understood". Another, more refined definition comes from Schwartz et al. (2002, 2007, in press) and can be formulated as follows:

> "Speech perception is the set of perceptual (auditory-visual; visual speechreading for the deaf; tactile TADOMA for the deaf-blind, Reed et al., 1982) processes allowing to recover and specify the timing and targets of speech gestures, supplying a set of representations for the control of one's own actions and the tracking and specification of somebody else's actions."

These are just two of the numerous and, in many cases, contradictory ways speech perception has been defined, which, for its part, implies that there is no consensus on this matter. Indeed, over the past 50 years, researchers have been trying to solve the question of how do humans perceive speech, but are still to achieve mutual understanding. The key question in speech perception research is how do listeners extract the significant information from the acoustic speech signal? How is the mapping between properties of the acoustic signal and linguistic elements, such as phonemes and syllables, done?

Distinct theoretical perspectives have been proposed to answer these questions. A crucial distinction among these perspectives can be put in the form of a question: what is the role of the speech motor system in speech perception? Or does it have a role at all? That is,

does speech perception rely exclusively on auditory perceptual mechanisms or is the perception mediated by the motor system (for reviews, see Schwartz et al. 2002, 2007; Diehl et al., 2004; Galantucci et al., 2006; Schwartz, 2008).

Many researchers advocate speech perception models that focus only on the auditory system and the acoustic properties of speech (for a review, see Diehl et al., 2004). This perspective comprises two important hypotheses: first, the same mechanisms are responsible for processing of both speech and nonspeech acoustic signals, and second, the objects of perception are the properties of the acoustic speech signal. However, a serious challenge for the latter hypothesis is posed by the ambiguity of the acoustic speech signal; for instance, the spectrum of the same phoneme within a speech signal varies depending on the preceding phoneme. This effect is due to phenomenon called coarticulation, which means that in speech the adjacent phonemes are not produced discretely one after the other, but instead with a partial temporal overlap.

In contrast to the hypotheses of the auditory perspective, the motor theory of speech perception (Liberman et al., 1967; Liberman and Mattingly, 1985; Liberman and Whalen, 2000) claims the objects of speech perception to be the speaker's intended articulatory gestures rather than the acoustic speech signal. Importantly, these intended gestures can be thought of as invariant, unlike the objects of perception claimed by the auditory approaches, thus making the motor theory more credible with regards to this hypothesis. Furthermore, the mechanisms proposed to be responsible for speech processing by the motor theory are thought to be special; that is, assuming that speech sounds are mapped into speech-specific representations, thus making processing of speech sounds radically different from that of non-speech sounds.

To put it plainly, whereas the motor theory asserts that perception is brought about by reconstruction of the motor movements that would have been necessary to produce the speech heard, the auditory theory, on the other hand, claims that ordinary auditory processing of the speech signal is sufficient to explain the perception of speech. Consequently, the motor theory assumes that there exists a very close relationship between speech production and perception; that is, the role of the speech motor system is not only to produce speech articulations but also to detect them.

Recently, several studies have argued against the purely auditory approaches, yet failed to give concurrent support the motor theory of speech perception, thus shining new light on the distinct bipartition of the two and sprouting new, interesting neurobiological models of speech perception. These recent findings have blown a fresh breeze over the decades-old debate on speech perception, making the question on the role of the speech motor system in speech perception, if possible, even more fascinating than it already was.

This thesis comprises a magnetoencephalographic study, within which the role of the speech motor system in speech perception was studied by investigating how two different speech stimulus types (syllables with different levels of ambiguity) presented in four different perception conditions (with varying motor tasks required from the subjects subsequent to the stimulus) affect the neural mechanisms underlying speech perception. The study of this thesis was carried out as a preliminary part of a larger project laying the groundwork for more refined and thorough analyses, within which the ideas of the recent neurobiological models are tested. Within the scope of this thesis, the principal aim was to test the hypotheses of the two distinct theoretical perspectives on speech perception by, for instance, observing the neural response changes with the two stimulus types (i.e., degradation of speech sounds) or whether the responses in the conditions with differing subsequent-to-stimuli motor tasks show a hierarchy of motor involvement from passive perception to the most active subsequent-to-stimuli motor task.

Chapter 2

Background

2.1 Magnetoencephalography

Transfer of information from one neuron to another in the brain involves electric currents. In accordance with Maxwell's equations, any electrical current will produce an orthogonally oriented magnetic field. Magnetoencephalography (MEG) is a totally non-invasive research method where brain function is studied by measuring these magnetic fields. It provides spatial resolution of 2-3 mm and can record the measured signals at an extremely high temporal resolution (on the order of 1 ms) giving it an edge over some of the other brain activity measurement techniques with much longer time scales, such as functional magnetic resonance imaging (fMRI). In this chapter, I will describe the origin and measuring of the neuromagnetic fields as well as the modeling of the measured activity, and lastly, briefly discuss MEG as a brain research tool and compare it to other imaging techniques. For more detailed description of the issues discussed in this chapter, see the extensive MEG review by Hämäläinen et al. (1993).

2.1.1 Origin of neuromagnetic fields

The human cortex contains about 10^{10} neurons (Williams and Herrup, 1988) connecting each other and working for information transferring. An individual neuron (Figure 2.1) consists of a cell body called soma, and extensions called dendrites and axons. Dendrites bring information to the soma and axons take information away from the soma to the next neuron. Information from one neuron to another flows across a synapse.

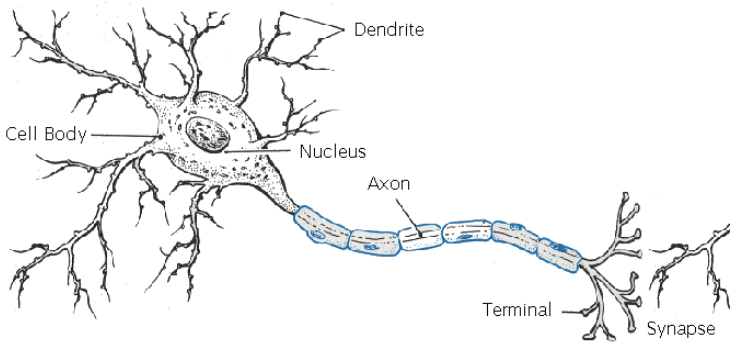

Figure 2.1: A neuron. Each neuron has three basic parts: cell body (soma), one or more dendrites, and a single axon. Adapted from Carlson (1992).

Within a neuron, the ability to transfer information is based on the functioning of its cell membrane. The membrane divides the cell into intra- and extracellular spaces with different ion concentrations. The concentration difference is maintained by proteins located in the membrane. These proteins act as gateways or ion pumps, transporting ions from regions of low concentration to regions of high concentration. The most important gateway is the Na-K -pump which transfers the Na^+ ions out and K^+ ions in. The concentration difference maintained by the ion pumps leads to diffusion as specific ion channels transport ions from regions of high concentration to regions of low concentration. This transport of ions through the ion channels then changes the voltage of the cell membrane and the resulting potential difference between the intra- and extracellular spaces opposes the flow of ions. The equilibrium or resting state is reached when these current flows created by diffusion and potential difference cancel each other. At that point, the voltage across the cell membrane is about -70 mV.

If the potential difference of the cell is decreased (i.e., the cell is polarized), voltage sensitive Na^+ -channels open, allowing Na^+ -ions to flow into the cell increasing depolarization further. The depolarization has an additional effect of opening up the nearby ion channels, and an action potential, a travelling constant amplitude voltage pulse, is created. The action potential lasts only a few milliseconds before the cell is repolarized as the voltage sensitive K^+ -channels open up.

The signal transferring from one neuron to another across a synapse usually happens with the help of chemical transmitters. When the action potential reaches the synaptic cleft, a tiny space between the pre- and postsynaptic cells, in the presynaptic cell, transmitter

molecules are released into the synaptic cleft. When these molecules reach the postsynaptic side, they affect the potential of the cell by either repolarization (inhibitory synapse) or depolarization (excitatory synapse). The net effect of all the synapses determine whether an action potential happens in the post-synaptic cell or not.

While the magnetic fields produced by action potentials are difficult to detect with MEG due to their electric properties, the post-synaptic potential generates a more easily detectable signal. Although individually the post-synaptic currents are weaker than action potentials, they can last for several tens of milliseconds which enables the temporal summation of their fields in parallelly oriented neurons. According to the right-hand rule, a current dipole gives rise to a magnetic field that flows around the axis of its vector component. Moreover, it has been shown that these magnetic fields are generated mainly in the pyramidal cells (Figure 2.2) that are oriented tangential to the surface of the cortex (Okada et al., 1997). Furthermore, it is often bundles of these neurons located in the sulci of the cortex with orientations parallel to the surface of the head that project measurable portions of their magnetic fields outside of the head (Figure 2.3).

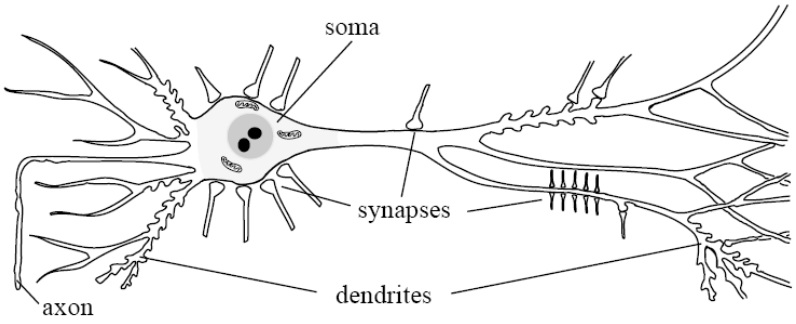

Figure 2.2: A pyramidal neuron. Information between neurons is transferred across synapses. If the net effect of the post-synaptic potentials depolarizes the cell enough, the neuron fires an action potential that travels along the axon. Modified from Iversen (1979).

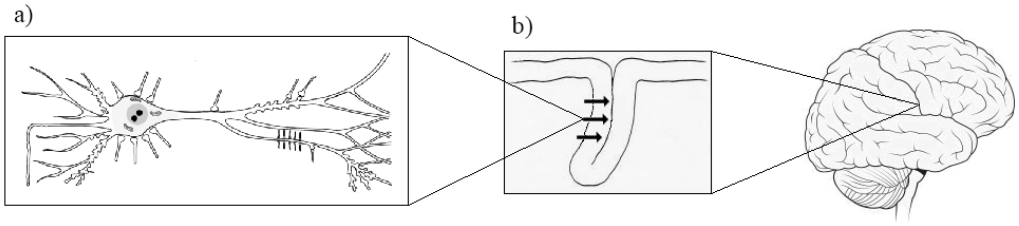

Figure 2.3: a) A pyramidal neuron. b) MEG is most sensitive to pyramidal neurons located in the sulci of the cortex with orientations tangential to the surface of the cortex and parallel to the surface of the head.

2.1.2 Measurement of neuromagnetic fields

The magnetic field generated by a single neuron is almost negligible, too small to be detected outside the head. Thus, it is only when several tens of thousands of nearby cells are synchronously active that they produce a magnetic field that can be measured outside the skull by the MEG (Hari, 1999). However, even when tens of thousands of neurons are synchronously active, the magnetic fields recorded outside the head are still extremely weak, typically between 50-500 fT (= 10^{-15} Tesla), which is several orders of magnitude smaller than the ambient magnetic noise in an urban environment, which is on the order of 10^{8} ft. Therefore, the weakness of the signal and strength of the competing environmental noise brings forth a problem akin to listening for the footsteps of an ant in the middle of a rock concert. To answer this problem, SQUID (Superconducting Quantum Interference Devices) sensors (Zimmerman and Silver, 1966), ultrasensitive detectors of magnetic fields, are used to measure the cortically generated signal. To reduce the level of the external magnetic noise, the measurements must be performed in a magnetically shielded room. Moreover, because the operation of SQUID is based on superconductivity, the sensors must be stored in liquid helium.

The devices used in MEG measurement are called neuromagnetometers. In a neuromagnetometer, a large number of sensors are laid in a grid formation that covers the entire scalp. A sensor typically consists of two parts, a SQUID and a flux transformer, which is further divided into a pick-up coil and a signal coil. The purpose of the flux transformer is to further enhance the sensitivity of the SQUID to magnetic fields. This is done by coupling the superconducting pick-up coil, having greater area and number of turns than the SQUID inductor alone, to the SQUID via the signal coil.

The properties of the measurement can be controlled by the design of the pick-up coil. A magnetometer (Figure 2.4a), a sensor with the pick-up coil containing only a single loop, measures the magnetic field directly at a given location and is thus more sensitive to magnetic fields from far away sources, which enables a better study of deep brain sources, but, on the other hand, exposes it to external magnetic disturbances. A different spatial sensitivity pattern can be achieved by the usage of a gradiometer, which consists of both a pick-up coil and an oppositely wound compensation coil in a planar or axial configuration. In planar gradiometers (Figure 2.4b), the pick-up and the compensation coil are placed next

to each other on the same plane, whereas in the axial configuration (Figure 2.4c) they are placed in series (i.e., on top of each other). This design measures the field difference between two locations and is, correspondingly, more sensitive to magnetic fields from nearby sources. The main benefit of the design is that a homogeneous field causes opposing flux through the oppositely wound coils thus providing effective noise cancellation as the practically homogenous fields generated by external magnetic noise sources (such as electric cables) induce equal but opposite currents in the loops. Therefore, gradiometers are more suitable for recording cortical activity than magnetometers when magnetic disturbances are present.

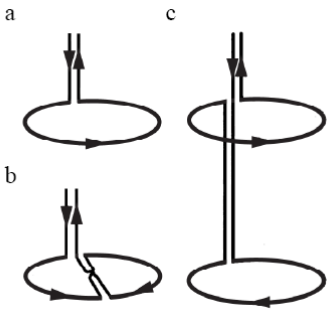

Figure 2.4: Three types of flux transformers: a) A magnetometer, b) a planar gradiometer, and c) an axial gradiometer. Adapted from Hämäläinen et al. (1993).

The study presented in this thesis was carried out with a 306-channel Vectorview™ device (Figure 2.5).

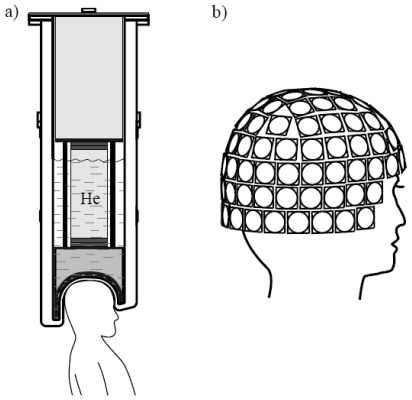

Figure 2.5: a) The Vectorview™ neuromagnetometer contains liquid helium to maintain the SQUID sensors in superconducting state. b) The Vectorview™ system has 102 recording sites, each with two orthogonal gradiometers and a magnetometer. Modified from Hämäläinen et al. (1993).

MEG recordings

In preparation for a MEG recording session, a small number of electrodes and coils are attached to the subject. The purpose of the electrodes is to monitor possible sources of muscle movements and artifacts, such as those caused by eye movements and blinks, so that the masking magnetic fields caused by them can be rejected from the MEG signal. The purpose of the coils, on the other hand, is to localize the head of the subject in the MEG device coordinate system (i.e., with respect to the SQUID sensor array). The coils should be placed as far from each other as possible, typically at four specific locations on the head: on the left and right mastoid, and on the left and right side of the forehead right below the hairline. The attached coils are then localized with respect to specific anatomical landmarks using a 3D digitizer. The landmarks commonly used are the left and right preauricular point and the nasion. These points also define the head coordinate system where the x-axis runs through the preauricular points, the y-axis is perpendicular to the x-axis and runs through the nasion, and the z-axis is perpendicular to both x- and y-axes (Figure 2.6).

After the preparation, the subject is seated under the helmet-shaped neuromagnetometer, the electrodes and coils are plugged in, and, in a typical MEG recording session, the subject is presented with a series of trials containing a stimulus (e.g., a spoken syllable) and a specific task related to it (e.g., repetition of the syllable). Due to low signal-to-noise ratio (SNR) of the MEG raw data, the trial is usually repeated 50–500 times during the experiment and the obtained data epochs (i.e., data sections within a trial) are typically averaged in order to improve the SNR.

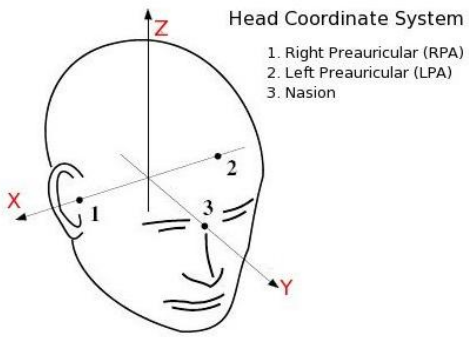

Figure 2.6: The landmarks and the head coordinate system used in determining head position in MEG recordings: the left and right preauricular point and the nasion.

2.1.3 Source modeling

In order to obtain an accurate spatiotemporal picture of the neural activation related to some specific sensory or cognitive operation performed by the subject, the inverse problem of neuromagnetism must be solved. The neuromagnetic inverse problem refers to the difficulty of locating the activated brain areas based on the measured magnetic fields as multiple different current distributions can generate the same fields outside the head (Helmholtz, 1853). Although unambiguous solution cannot be achieved, source models based on the physiology of the brain can be used to address this problem.

The most commonly used source model representing cortical activity is the dipole model (Kaufman et al., 1981; Tuomisto et al., 1983). In this model, the current source is modeled as a point-like equivalent current dipole (ECD) with specific location, orientation, and strength. A set of ECDs are calculated from the recorded MEG data by minimizing the difference between a calculated and the measured magnetic fields using least-squares search.

Multiple other source modeling techniques are also available for solving the MEG inverse problem. These include minimum norm estimates (Hämäläinen and Ilmoniemi, 1984; Uutela et al., 1999; Jensen and Vanni, 2002), beamforming (Van Veen and Buckley, 1988; Robinson and Vrba, 1997; Van Veen et al., 1997; Gross and Ioannides, 1999; Sekihara et al., 2001), current multipoles (Jerbi et al., 2004), and others ((Mosher et al., 1992; Baillet et al., 1999; Mosher and Leahy, 1999; Baillet et al., 2001; Gavit et al., 2001; David et al., 2002; Grasman et al., 2004).

2.1.4 Comparison with other brain activity measurement techniques

At present, a number of non-invasive brain activity measurement techniques are available to investigate the question of how and where in the brain particular sensory and cognitive processes occur. Instead of being competing, different techniques, such as positron emission tomography (PET), electroencephalography (EEG), functional magnetic resonance imaging (fMRI), and MEG, each provide strengths that can be complementary to each

other, thus opening new non-invasive windows through the human skull to the functioning of the brain.

Since MEG takes its measurements directly from the activity of the neurons themselves, it provides extremely high temporal resolution, that of a millisecond time-scale, thus allowing practically real-time brain activity monitoring. This is the main advantage of MEG, making it ideal for measuring the rapid changes in brain activation also during speech perception, given that both the perceptually significant alterations in acoustic speech signals and the following cognitive processes in the brain take place on a timescale of milliseconds. The spatial aspect, however, in the form of source localization is the most problematic feature of MEG due to the non-uniqueness of the inverse problem.

Both PET and fMRI, on the other hand, do not suffer from the non-uniqueness of the inverse problem, thus providing more accurate spatial resolution. However, in contrast to MEG, they have limited temporal resolutions mainly due to the larger timescale (hundreds of milliseconds) of the hemodynamic response (i.e., changes in blood oxygenation and flow in response to neural activity) which is what these methods exploit.

Although EEG and MEG are generated by the same neurophysiologic processes, there are some notable differences. Firstly, magnetic fields, in contrast to electric fields, are less distorted by the resistive properties of the skull and scalp, which result in a better spatial resolution of the MEG. Secondly, EEG detects activity from neural currents in any orientation, while MEG is most sensitive to tangential components of currents. In other words, MEG selectively measures the activity in the sulci, while EEG measures activity both in the sulci and at the top of the cortical gyri.

These attributional differences show, for example, how MEG and fMRI or MEG and EEG can be combinedly used to solve the inverse problem and thus achieve better source localization.

2.1.5 MEG as a research tool

MEG can be used both as a research tool to investigate several sensory and cognitive processes and in clinical use to study, for example, epilepsy. In addition to being evidently suitable for studies of the basic sensory functions, such as vision (Brenner et al., 1975), somatosensation (Brenner et al., 1978), and audition (Hari et al., 1980), it has also been substantial in studies of cognitive processes, such as those involved with language processing, including audiovisual speech perception (Sams et al., 1991) and speech production (Salmelin et al., 2000).

Although the main clinical uses of MEG are in epilepsy diagnosis and treatment, it can be also used in surgical planning to provide critical functional information by mapping out sensory-motor cortex (Ganslandt et al., 1996; Pang et al., 2008). Other clinical applications include, for instance, investigations of neurological disorders such as Alzheimer's (Stam et al., 2008) and Parkinson's disease (Stoffers et al., 2008).

2.2 The human auditory system

The human auditory system is the sensory system for the sense of hearing. Its purpose is to convert changes in the air pressure into neural impulses that are then perceived and interpreted by the brain as sound. For a detailed review of the human auditory system and its functions, see e.g. Goldstein (2007).

2.2.1 Structure of the human ear

Sound enters the auditory system through the outer ear (Figure 2.7), is transformed to a suitable range of pressure changes by the middle ear, and finally, is converted to neural signals inside the inner ear, in the cochlea.

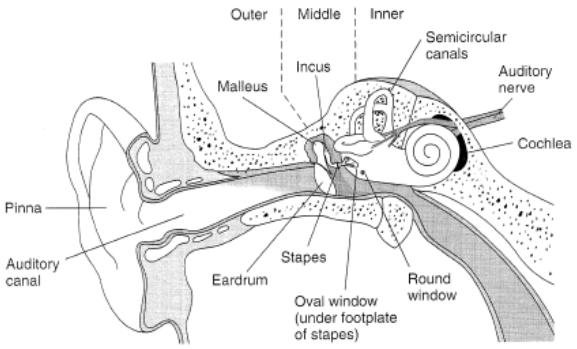

Figure 2.7: The structure of the ear. Adapted from Goldstein (2007).

Outer ear

The outer consists of the *pinna* and *auditory canal*. The pinna comprises the folds of cartilage surrounding the ear canal. It reflects and attenuates arriving sound waves and, due to its structure, helps the brain to determine the direction from which the sound came. The auditory canal is about 3 cm long, wax-covered tube that both protects the middle ear from the outside world and amplifies sounds in the range 3 to 12 kHz. At the end of the auditory canal is the *tympanic membrane* (or *eardrum*), which marks the beginning of the

middle ear. Any vibrations of the tympanic membrane caused by the sound wave are transferred to the middle ear.

Middle ear

The middle ear is an air-filled cavity made up of a series of delicate bones known as the *ossicles*. These three minuscule bones are called the malleus (hammer), incurs (anvil), and stapes (stirrup). As the sound waves travel through the auditory canal, they hit the tympanic membrane causing it to vibrate, which, in turn, causes the malleus, attached to the tympanic membrane, to vibrate. Malleus passes the vibrations on to the second bone, incus, and the incus to the last bone, stapes.

The ossicles transfer the lower-pressure vibration of the tympanic membrane into higher-pressure sound vibrations at another membrane-covered opening called the *oval* (or *elliptical*) *window* (attached to the stapes), marking the beginning of the inner ear. Higher pressure is necessary because the inner ear beyond the oval window contains liquid rather than air. The ossicles, thus, act as an impedance matching device, ensuring the easy and efficient propagation of sound energy to the inner ear fluids.

Inner ear

The main structure of the inner ear is the *cochlea* (Figure 2.8 and 2.9). Cochlea is a coiled tube filled with fluid, roughly 2 mm in diameter and 3 cm in length. It is divided into three chambers along its length. Two of the chambers called *scala media* (or *cochlear duct*) and *scala tympani* are separated by the *basilar membrane*. The third chamber, called *scala vestibule*, is correspondingly separated from the scala media by *vestibular* (or *Reissner's*) *membrane*. Scala media is filled with extracellular fluid and, sitting above the basilar membrane, contains the *organ of Corti*.

The organ of Corti supports about 30000 hair cells with nerves connected to each of them. The hair cells are arranged in a row of inner cells and three rows of outer cells. Each hair cell contains a bundle of 100-200 specialized *cilia* (or hair) at the top, acting as the sensors

for any mechanical vibrations. The tectorial membrane rests above the longest cilia and moves back and forth with each cycle of sound. The movement of tectorial membrane tilts the cilia and allows electric current into the hair cell. The *vestibulocochlear nerve* attached to the hair cells fires and transmits the message to the auditory region of the brain. In other words, the hair cells work as mechanotransducers, converting mechanical energy into neural signals to the brain.

The basilar membrane is stiffest near the oval window, and becomes more flexible toward the opposite end, allowing it to act as a frequency spectrum analyzer. Hair cells close to the oval window transmit information about high-frequency sound, while those at the far end of basilar membrane provide information about low-frequency sound.

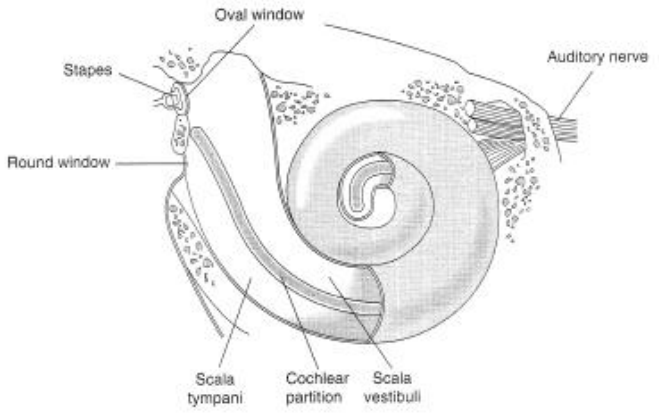

Figure 2.8: Cochlea. Adapted from Goldstein (2007).

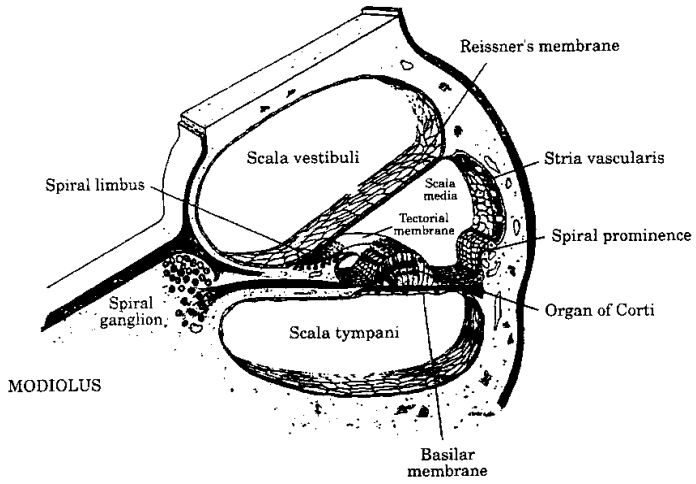

Figure 2.9: Cross-section of the cochlea. Adapted from Bloom and Fawcett (1975).

2.2.2 Auditory pathway

The auditory pathway starts from the cochlea (Figure 2.10), from where the re-encoded sound information travels down the vestibulocochlear nerve, through the intermediate stations of the *cochlear nucleus* and the *superior olivary nucleus* in the brain stem, and the *inferior colliculus* of the midbrain, being further processed at each waypoint. The information eventually reaches the *medial geniculate nucleus* of the thalamus (MGN), and from there it is relayed to the *primary auditory cortex* (PAC, or in non-human primates also called A1), located in the temporal lobe of the cortex.

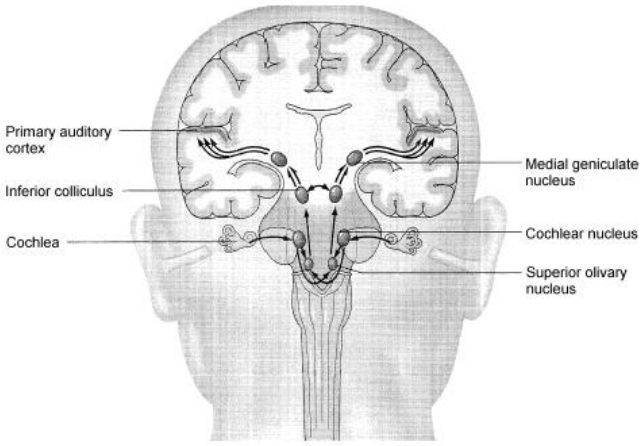

Figure 2.10: The auditory pathway. Adapted from Goldstein (2007).

The auditory pathway is bilaterally organized, meaning that they exist in both sides of the brain (see Figure 2.10). The pathway is called ipsilateral if the auditory input comes from the cochlea in the same side and contralateral if it comes from the opposite side of the auditory cortex.

2.2.3 Auditory cortex

Auditory cortex (AC) is the region of the cerebral cortex that is responsible for processing of auditory information. It is located on the temporal lobe and is concentrically organized in different areas (a primary area and several peripheral, or belt, areas) with the primary area (i.e., PAC) in the middle. PAC is located in the *superior temporal gyrus* (STG) in the depth of

the Sylvian fissure (also called *lateral sulcus* or *lateral fissure*) where it occupies the medial two-thirds of *Heschl's gyrus* (HG, also called *transverse temporal gyrus*) and corresponds closely to Brodmann areas (BA) 41 and 42 (Figure 2.11 and 2.12).

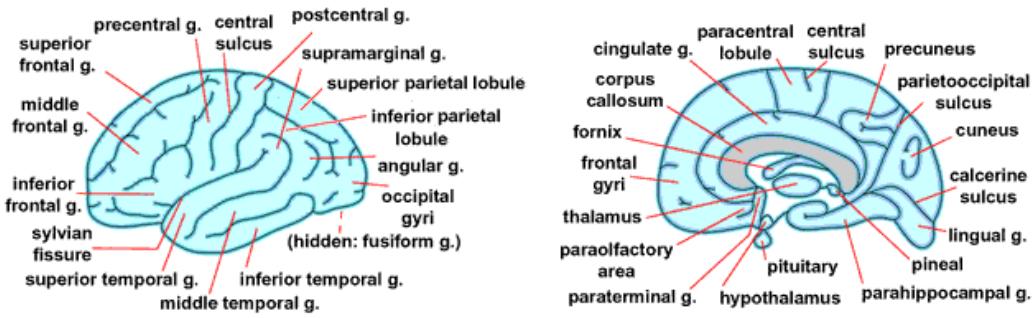

Figure 2.11: Diagram of the main gyri and sulci of the brain. Modified from Dubin (2009).

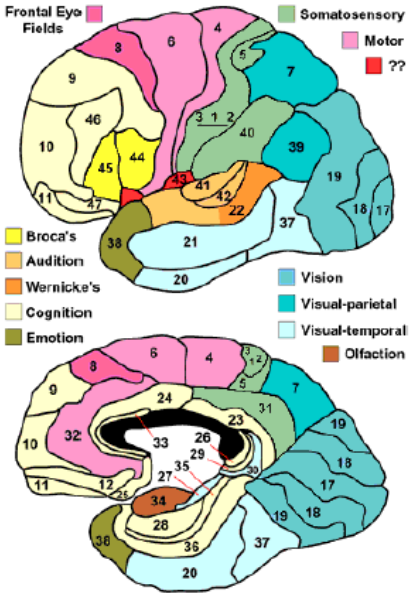

Figure 2.12: Diagram of functional areas of the brain, depicting also the corresponding Brodmann areas. Adapted from Dubin (2009).

When the auditory input from MGN of the thalamus reaches AC, it is first received in the core area, including PAC and some nearby areas (Figure2.13). Signals are then sent to areas surrounding the core, called the *secondary auditory cortex* and after that, to the *auditory association cortex* (Goldstein, 2007).

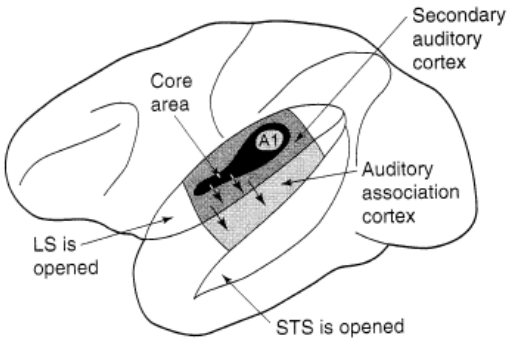

Figure 2.13: Organization of the auditory cortex. LS stands for lateral sulcus (i.e., Sylvian fissure) and STS for superior temporal sulcus. In this picture both the LS and STS are ”opened” to expose part of the auditory cortex not visible from the surface. The figure is actually of the left hemisphere of a macaque monkey brain, but applies for illustrating the organization of human auditory cortex as well. Adapted from Goldstein (2007).

PAC is tonotopically organized, meaning that its neurons are frequency specific, organized according to the frequency to which they respond best. Figure 2.14 illustrates the tonotopic organization with the characteristic frequencies.

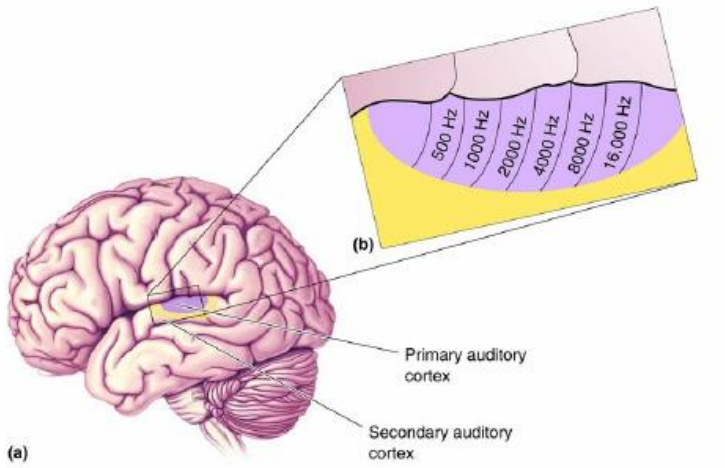

Figure 2.14: Primary auditory cortex (purple) and secondary auditory cortex (yellow) on the superior temporal lobe. (b) Tonotopic organization in primary auditory cortex with characteristic frequencies depicted in Hertz. Adapted from Bear et al. (2001).

2.3 Event-related fields

A brain response measured with MEG that is directly the result of an external stimulus is called an *event-related field* or an *evoked response field* (ERF) and is the equivalent to the electroencephalographically measured *event-related potential* (ERP). The brain response to a single stimulus, however, is usually impossible to be visually differentiated from the raw MEG signal because of the thousands of other simultaneously ongoing brain processes. To see the brain response to the stimulus, many (usually 100 or more) trials must be conducted, within which short segments (epochs) of data time-locked to the stimulus are averaged together. The averaging causes all the random and irrelevant brain activity to be canceled out and the relevant ERF to remain, resulting in improved signal-to-noise ratio (SNR). The more averaged epochs, the better the SNR will be. The averaging process is illustrated in Figure 2.15.

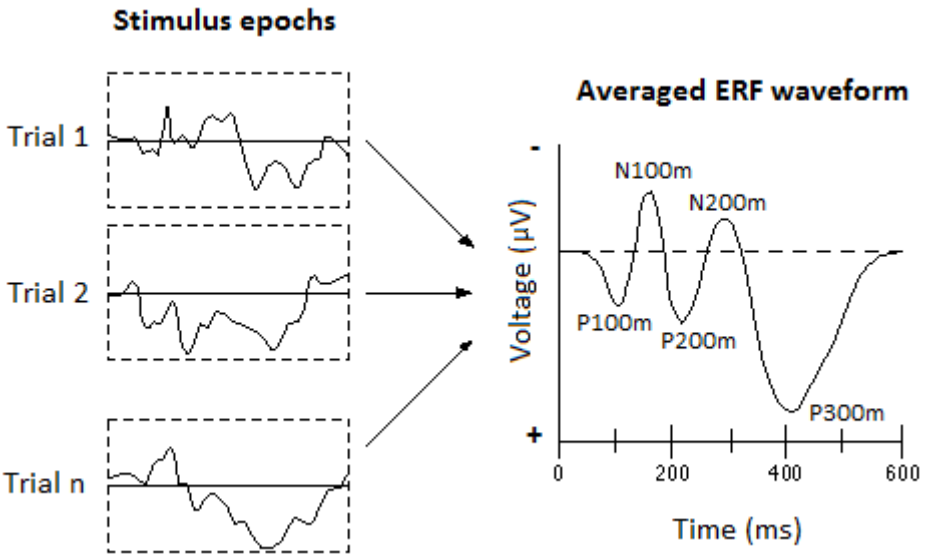

Figure 2.15: The ERF averaging process. Averaging the raw MEG signal filters out brain activity that is irrelevant to stimulus presentation. The left side of the figure depicts raw data segments phase-locked to stimulus onset, while the right side shows the averaged ERF waveform and its components.

ERFs/ERPs can be elicited by a number of different stimuli, such as a flash of light or a speech sound. The ERF/ERP elicited by the former is called a *visual evoked field/potential* (VEF/VEP), while that elicited by the latter is known as an *auditory evoked field/potential* (AEF/AEP). The next chapter focuses on AEFs.

2.3.1 Auditory evoked fields

The neural responses to auditory stimuli can be divided into three categories: early, middle, and late latency components (Picton et al., 1974). The early latency AEFs occur within the first 12 ms following the sound stimulus onset and originate in the cochlea (within the first 4 ms) and brain stem (within the first 12 ms). The middle latency components occur between 12-50 ms after the stimulus onset and are believed to originate in or near the primary auditory cortex (Mäkelä et al., 1994; McEvoy et al., 1994; Yoshiura et al., 1995). The late components refer to any evoked fields occurring after 50 ms from the stimulus onset and are generated within the primary and secondary auditory cortices (e.g., Scherg et al., 1989).

The late AEFs can be further divided into components (see Figure 2.15). For example, N100m refers to the negative peak at around 100 ms and, correspondingly, P200m to the positive peak at around 200 ms from the stimulus onset. The letter *m* in the component names is just to clarify that the response is obtained using MEG. Other than that, the components are identical to the EEG equivalent, which are denoted without the additional letter (e.g., N100).

N100m is usually the most prominent peak of the AEF and is also the subject of interest in this thesis. The earlier late latency components, such as N100m and P200m are primarily dependent on the characteristics of the stimulus, whereas the later late components, such as P300m and N400m, are more dependent on internal cognitive processes (Celesia and Brigell, 1999).

Auditory N100(m)

The auditory N100 is generated by a network of neural populations in the primary and association auditory cortices in the superior temporal gyrus in Heschl's gyrus (Zouridakis et al., 1998) and planum temporale (Godey et al., 2001), located just posterior to HG, within the Sylvian fissure. As implied before, its amplitude (and latency) is strongly dependent

upon the stimulus characteristics, such as the rise time of the onset of a sound (voice onset time, VOT; Spreng et al., 1980), loudness (Keidel et al., 1965), and interstimulus interval (ISI; Davis et al., 1966). Regarding ISI, the discovery has been made that the amplitude of N100 shows deflection upon repetition of a stimulus; in other words, it first decreases with repeated presentations of the stimulus, but after a short period of silence it returns back to its previous level (Näätänen and Picton, 1987).

The N100 also depends on the unpredictably of stimulus, being weaker when stimuli are repetitive, and stronger when they are random. Furthermore, it has been shown that the amplitude of the component is reduced when subjects are warned about an upcoming stimulus (Schafer et al., 1981) and may even disappear when they are allowed to control the onset of stimuli (Schafer et al., 1973). In addition, it has been suggested that the stimulation resulting from efference copies from the intended movements of a person is not processed (Kudo et al., 2004), causing, for example, a reduced N100 component to be produced by a person's own voice (Curio et al., 2000).

2.4 Neural basis of speech production

Over a century ago, a French neurologist Paul Broca demonstrated, by examining the brains of aphasic patients, that speech mechanisms could be localized in the human brain. He discovered that the patients' inability to speak was due to a lesion in the inferior part of the frontal lobe (Broca, 1861a, 1861b). However, since the time of Broca, scientists have found that lesions to this area, aptly named Broca's area (see Figure 2.12), alone are not enough to produce lasting speech deficits (e.g., Alexander et al., 1989; Dronkers et al. 2000; Mohr 1976). This section, based mainly on Dronkers and Baldo (2001), focuses on the brain mechanisms and neural structures involved in speech production from a conception of an utterance to its ultimate articulation.

Numerous studies have confirmed the importance of the temporal lobe on the translation of concepts into linguistic representations. Once the utterance has been linguistically formulated, it must be transferred to the speech mechanisms that will execute its production. Articulation itself requires planning, initiation, modification, and execution. It

has become evident that many areas, along with Broca's area, are involved in these aspects of speech production. Recent research has suggested a role for the insular cortex in articulatory planning (Dronkers, 1996); the supplementary motor area for the initiation of sequential speech movements (Ziegler et al., 1997); the basal ganglia and the cerebellum in the modification of pitch, loudness and rate (e.g., Duffy, 1995; Fabbro et al., 1996; Coplin et al., 1997; Pickett et al., 1998); and the primary motor face cortex and pre-motor cortex in the execution of articulatory movements (e.g., Wildgruber et al., 1996; Ackermann et al., 1998).

Arcuate fasciculus

The *arcuate fasciculus* is the neural pathway connecting the temporal and frontal lobes and is considered as part of the *superior longitudinal fasciculus*. It was identified by Norman Geschwind as the tract that could potentially connect Wernicke's language area to Broca's speech area as described in Wernicke's original model (Geschwind 1970; Wernicke 1874; Figure 2.16).

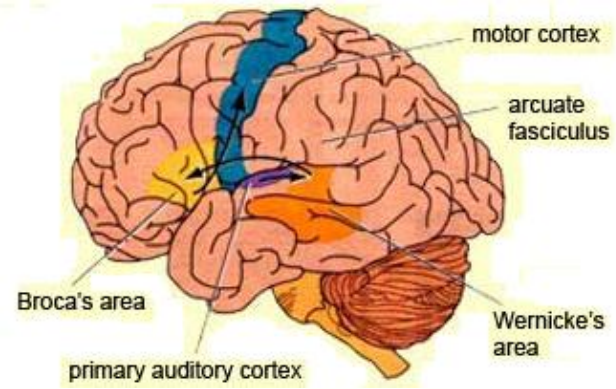

Figure 2.16: The Geschwind model depicting the brain areas associated with language comprehension and production. The figure illustrates the brain mechanisms and neural structures involved in pronouncing a word after hearing it. Adapted from Dubuc (2010).

The Geschwind model predicts that disruption of this tract results in isolated repetition deficits. However, recently Dronkers et al. (2000) reported a more comprehensive disorder in that lesions that sever the arcuate / superior longitudinal fasciculus result in a severe production deficit with a complete loss of coherent, propositional speech, leaving only

automatized utterances to remain. These patients have lost more than just repetition skills; that is, they are unable to transfer any information from the temporal lobe language areas to the anterior speech mechanisms.

Insular cortex

The insular cortex (often called *insula*) is a region of neocortex deep within the Sylvian fissure in the intersection of the frontal, parietal, and temporal lobes. Whereas in the past the role of the insula in speech production was somewhat vague, more recently lesions in a specific region of the insula has been linked to an inability to plan and coordinate the appropriate movements necessary for articulation (Dronkers, 1996). Furthermore, recent neuroimaging studies have reported activation in the insula with tasks such as articulation of single words, word reading, picture naming, and word generation (Indefrey and Levelt, 2000). Interestingly, the insula appears to be activated only when tasks involve articulation of non-repeated and phonologically complex words (e.g., Wise et al., 1999), but not when only automatic or simple articulatory patterns are produced (e.g. Murphy et al., 1997).

Supplementary motor cortex

The supplementary motor area is located in the *superior frontal gyrus* (see Figure 2.11) and is a part of the sensorimotor cerebral cortex. In a study investigating the participation of the supplementary motor area in speech, Penfield and Roberts (1959) showed that electrocortical stimulation of the area caused involuntary vocalizations or interruptions in speaking ability. A more recent review of the lesion literature suggests that the supplementary motor area is involved in the initiation of sequential, voluntary movements, including those for speech (e.g., Ziegler et al., 1997). Other tasks that have been shown to elicit supplementary motor cortex activation in neuroimaging studies include the control of breathing for speech and vocalization (Murphy et al., 1997) and automatic speech (e.g., reciting months of the year; Ackermann et al., 1998).

Basal ganglia

The basal ganglia (or *basal nuclei*) are subcortical nuclei that interact with the cortex and with a number of other subcortical structures in a series of feedback loops that help to maintain motor activity (see Love and Webb 1996). Lesions to the basal ganglia can result in several types of motor disorders including Parkinson's and Huntington's disease, each with their own characteristic speech disorders (Duffy 1995). For instance, patients with Parkinson's disease show hypokinetic speech with decreased intensity and little modulation of pitch or loudness. Lesions in the basal ganglia have been demonstrated to result in dysarthric (including disorders in many of the speech subsystems, such as respiration and prosody) and incomprehensible speech, with poor articulatory and phonatory control (Pickett et al., 1998), and, similarly, in hypophonia and a reduction of speech initiation and output (Fabbro et al. 1996).

Cerebellum

The cerebellum is a region of the brain that plays an important role in motor control in that it coordinates sensory input with muscular responses. Anatomically it has the appearance of a separate structure located just below and behind the cerebral hemispheres and above the brain stem (Figure 2.17).

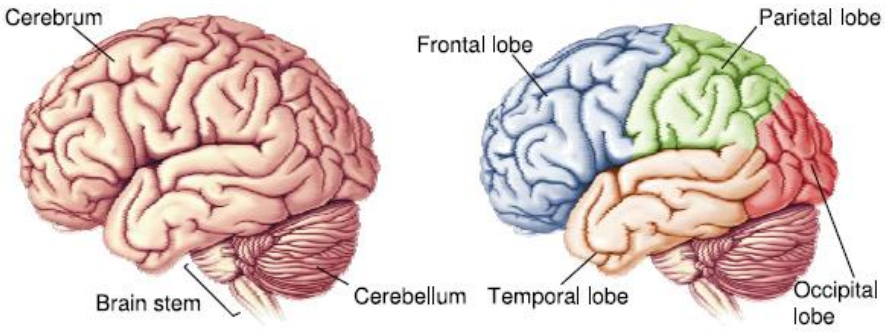

Figure 2.17: Anatomy of the cortex, depicting the main parts of the brain and the lobes of the cortex. Adapted from Bear et al. (2001).

The cerebellum has traditionally been thought to be involved in fine coordination of motor acts including speech. A broader view suggests, however, that the cerebellum acts as a timing mechanism, that is, the role of the cerebellum in motor coordination (for speech and other coordinated motor acts) is based on a more general mechanism allowing the precise temporal control of such movements (Ivry and Keele, 1989). Consistent with this hypothesis, several studies have shown that cerebellar patients have abnormal voice onset time distributions, indicating a timing disorder (e.g., Gandour and Dardarananda, 1984). In related, more recent studies, the cerebellum was suggested to function as a store for verbal short-term memory with respect to speech production (Silveri et al., 1998; Ivry and Fiez, 2000).

Broca's area

Broca's area is now typically defined in terms of the orbital and triangular parts (*pars opercularis* and *pars triangularis*) of the inferior frontal gyrus (IFG), encompassing Brodmann areas (BA) 44 and 45 (Dronkers et al., 2007; see Figure 2.12). The traditional view has emphasized Broca's area as a major structure for speech production. Recently, however, the function of the area has been re-evaluated. Indeed, it has been shown that lesion (or resection) of Broca's area results only in a transient mutism that resolves in 3-6 weeks (Dronkers et al., 2000; Penfield and Roberts, 1959). In their study, Dronkers et al. (2000) also demonstrated that focal lesions to Broca's area do not result in a persisting Broca's aphasia, and that Broca's aphasia (i.e., inability to produce language [spoken or written]) may result from lesions outside of Broca's area.

With the increasing understanding on the contribution of other brain areas to speech production, Broca's area is no longer considered to be as crucial and comprehensive with respect to speech production as once thought. The prevalent conception of Broca's area is that, although clearly involved in articulation, it operates within a network of brain regions that support speech production. Indeed, neuroimaging studies have suggested networks of activation that include Broca's area in tasks involving phonological encoding and articulation, such as picture naming, word generation, and reading (e.g., Indefrey and Levelt, 2000).

Motor face cortex and pre-motor cortex

The primary motor cortex (M1) is located in the dorsal part of the precentral gyrus and the anterior bank of the central sulcus (see Figures 2.11 and 2.12). Figure 2.18 illustrates which part of the human body is controlled by which part of the M1. Like most brain functions, motor controls are also crossed; that is, the right motor cortex controls the left side of the body, and the left motor cortex controls the right side. The pre-motor cortex is located just anterior to the M1, and is largely equivalent to BA 6.

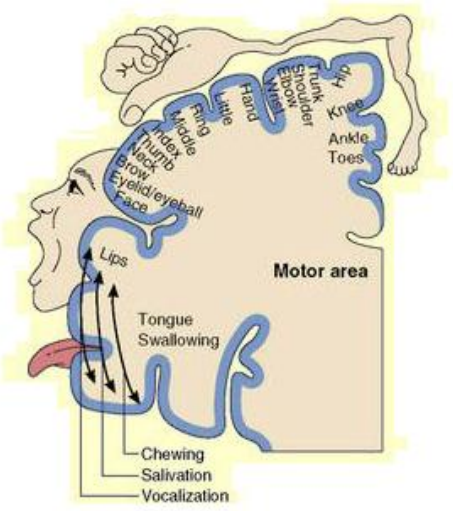

Figure 2.18: The motor homunculus, i.e., a map illustrating which part of the M1 controls which part of the human body. Adapted from Dubuc (2010).

The motor face area of primary motor cortex and the pre-motor cortex are believed to be the source the impulses that stimulate muscles of the vocal mechanism (via cranial nerves V, VII, IX, X, XI, and XII; Duffy, 1995). Lesions to these areas result in a pure motor speech disorder with slow, effortful speech and impaired articulation yet intact language (Alexander et al. 1989). Studies employing electrocortical stimulation have showed that a number of regions in primary motor and pre-motor cortex are involved in speech output, as stimulation of these sites resulted in speech arrest (Penfield and Roberts, 1959; Ojemann and Mateer, 1979). More recent fMRI studies of speech production have found that, for instance, that tongue movements elicited comparable right and left activation in the inferior region of the motor strip (Wildgruber et al., 1996), and that monotone overt as well as covert recitation of the months of the year activated more left than right motor cortex while singing activated more right than left motor cortex (Ackermann et al., 1998).

2.5 Speech perception

For most of us, understanding speech is a relatively effortless task. We are able to comprehend speakers in noisy environments and cope with a vast variety of different pitches, intonations and dialects, all of which affect the properties of the speech signal we need to decode in order to understand what was said.

Over the past 50 years, there has been one question above all others puzzling the minds of researchers in speech perception; that is, how do listeners extract the significant information from the acoustic speech signal? How is the mapping between properties of the acoustic signal and linguistic elements, such as phonemes and syllables, done?

Although this mapping has proved to be rather complex and the answer to the key question still remains somewhat elusive, several distinct theoretical perspectives have been proposed to explain speech perception. These theories can be roughly categorized with respect to two factors (Diehl et al., 2004): (1) speciality of the mechanisms with which speech sounds are processed and (2) the proposed objects of speech perception. Table 2.1 illustrates this classification and lists the main theoretical approaches following the categories.

Theories claiming special mechanisms in speech sound processing (first column of Table 2.1) assume that speech sounds are mapped into speech-specific representations thus differentiating the processing of speech sounds from that of non-speech sounds (Liberman et al., 1967; Liberman and Mattingly, 1985). General mechanism theories (second column of Table 2.1), on the other hand, assume the same mechanisms for the processing of both speech and non-speech acoustic signals (Fowler, 1996; Massaro, 1998). Moreover, the theoretical approaches on the first row of Table 2.1 assume that the objects of speech perception are the articulatory gestures of the talker (Liberman et al., 1967; Liberman and Mattingly, 1985), whereas the ones on the second row assume the objects to be the acoustic speech signals (Diehl and Kluender, 1989; Massaro, 1998; Kuhl, 2000).

Table 2.1: The main theoretical approaches to speech perception. Adapted from Diehl et al. (2004).

	Special mechanisms	**General mechanisms**
Gestural objects	Motor theory	Direct realist theory
Non-gestural objects	Eclectic specializations	General approach

2.5.1 Theories on speech perception

This section is largely based on Diehl et al. (2004).

Motor theory of speech perception

When we hear spoken words we know that they are made of auditory sounds. The motor theory of speech perception argues, however, that behind the sounds we hear are the intended movements of the vocal tract that pronounces them (Liberman et al., 1967; Liberman and Mattingly, 1985). In other words, the motor theory (MT) claims that the objects of speech perception are articulatory events rather than acoustic or auditory events. Furthermore, it is important to emphasize that it is the neuromotor commands to the articulators (e.g., tongue, lips, and vocal folds), also referred to as the intended articulatory gestures, rather than actual articulatory movements or gestures that make the basis of speech perception according to MT.

Consequently, MT assumes that there exists a very close relationship between speech production and perception; the role of the speech motor system is not only to produce speech articulations but also to detect them. In other words, speech inputs are mapped to the same motor programs that observers use also in their own speech production. That is to say, understanding the speaker takes place when the articulatory gestural representations of the listener are activated by the listening to verbal sounds. This theory is supported by one of the classic problems in speech perception research, that is, the phonetic categories do not strictly correspond to the acoustic properties of phonemes because of coarticulation (e.g., phoneme /d/ is acoustically very different in syllables /di/

and /du/). Figure 2.19 illustrates this problem of mapping between phonemes and their acoustic realizations.

Another important claim made by MT is that speech perception is done through a specialized module that is speech-specific, innate, and unique to humans. In other words, the human ability to perceive speech cannot be ascribed to general mechanisms of audition and perceptual learning but instead depends on a specialized speech decoder. The decoder was hypothesized by Liberman et al. (1967) to operate by "somehow running the process of speech production backward" using an analysis-by-synthesis approach in which invariant gestural representations are at the origin of the acoustic signal, in order to recover the intended articulatory gestures of the speaker. As such, perceiving speech is perceiving speech motor (articulatory) gestures.

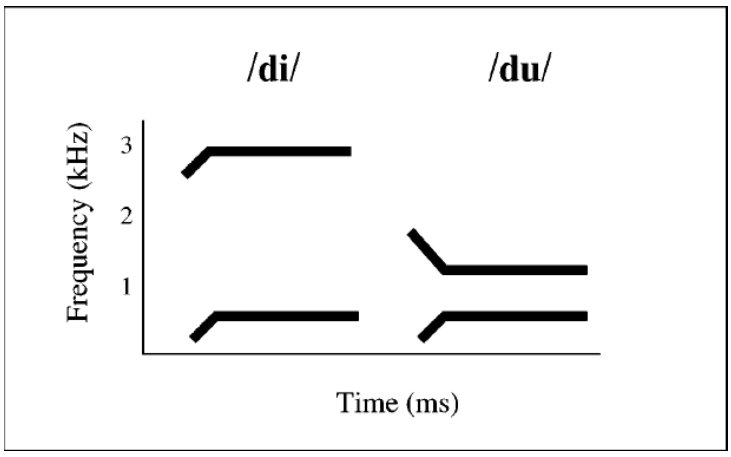

Figure 2.19: Simplified formant patterns for the syllables /di/ and /du/. What is noteworthy is that the second-formant transitions are not the same, even though the tongue tip gestures and the perceptions of /d/ are the same in both syllables. This difference arises from the fact that adjacent vowels and consonants are coarticulated (i.e., produced with temporal overlap). Adapted from Delattre et al. (1952).

Direct realist theory of speech perception

As Table 2.1 expresses, direct realist theory of speech perception (DRT), like MT, claims that the objects of speech perception are articulatory rather than acoustic signals (Fowler, 1986, 1996). However, while MT asserts the objects to be events that are causally antecedent to articulatory movements, such as neuromotor commands or intended gestures, according to

DRT, the objects are the actual movements or gestures themselves. Furthermore, in contrast to MT, DRT denies the speciality of the mechanisms with which speech sounds are processed and instead argues that same mechanisms are employed for the processing of all acoustic signals.

In fact, the theory argues that perceiving gestures, not the acoustic signal, in speech perception is analogous to perceiving a surface, not visible light, in visual perception. In other words, one directly perceives the event in the environment that has caused the structure in the media (e.g., acoustic signal), not the media itself. To sum up, DRT claims that the gestures of a talker (e.g., the closing and opening of the lips during the production of /pa/), structure the acoustic signal (by changing the shape of the vocal tract), which then serves as the informational medium that allows the listener to recover the gestures.

General auditory and learning approaches to speech perception

Several models of speech perception that focus only on the auditory system and the acoustic properties of speech have been proposed as alternatives to both MT and DRT (Diehl and Kluender, 1989; Massaro, 1998; Kuhl, 2000). More specifically, these models claim that speech is processed by the same mechanisms as non-speech sounds and that the recovery of the spoken message within the acoustic signal is not mediated by the perception of articulatory gestures. The term “general approach” (GA) is commonly used to depict such models (see Table 2.1).

A noticeable finding supporting this approach was that both speech and non-speech sounds can be perceived categorically (Stevens and Klatt, 1974; Miller, 1976; Pisoni, 1977). These similarities in perceptual performance, together with further findings showing that even non-human animals are able to perceive speech sounds categorically (something that was assumed by motor theorists to be unique to humans; Kuhl and Miller, 1975, 1978), lead to the claim that, instead of specialized mechanisms and/or gestural objects, speech perception relies on general auditory mechanisms with acoustic events being the objects of speech perception. An important factor contributing to GA is a process called perceptual learning, which is thought to enable us to learn the perceptual characteristics of our native language by detecting patterns and extracting statistical information from our auditory

environment during early development (Kuhl, 2000). GA assumes therefore that speech sounds are perceived using the same mechanisms of audition and perceptual learning that were evolved in humans (or human ancestors) to handle other classes of environmental sounds.

The term “eclectic specializations” on the lower left corner of the Table 2.1 is reserved for theories claiming that speech perception uses special mechanisms to recover a non-gestural representation of linguistic elements. In my best knowledge, no such (coherent) theories has yet been established; however, several proposals that special mechanisms of speech sound processing may work in concert with general perceptual mechanisms. One such proposal, attributed to specialized processes of categorization, is the finding that human infants possess an ability to learn the phoneme categories of their native language (Kuhl, 1993).

2.5.2 Cortical organization of speech perception

In accordance to the juxtaposition of auditory vs. motor theories of speech perception, the key question considering the cortical processing of speech is whether acoustic speech signals are processed by speech-specific neural mechanisms or by the same auditory mechanisms as other complex sounds. Furthermore, if the former is true, at which processing stage do the specialized mechanisms exist?

To answer these questions, a number of neuroimaging studies comparing responses to both speech and non-speech sounds have been conducted over the years (Binder et al., 2000; Scott et al., 2000; Vouloumanos et al., 2001; Narain et al., 2003). The results have been consistently showing that the activation for speech sounds in the left superior temporal regions (i.e., *superior temporal gyrus* [STG] and *superior temporal sulcus* [STS]) is prominently greater than that of non-speech sounds. Processing of non-speech sounds, on the other hand, has been attributed to the *primary auditory cortex* (PAC) and dorsolateral portions of STG. These findings suggest that whereas the processing of the acoustic features of non-speech sounds occurs at early levels in the PAC (in BA 41), the phonetic processing of speech sounds occurs at later level in the auditory processing stream in the

secondary auditory cortex (in the surrounding regions of STG and STS encompassing BAs 42, 21, and 22; Rademacher et al., 1993).

However, it is rather debatable to draw conclusions about speech-specific mechanisms from comparing cortical activations evoked by speech and non-speech sounds that are acoustically distinctly different. That is, any observed differences in response may simply be due to differences in the acoustic features of the signals rather than the distinction of whether a signal is speech or not. Thus, it may be that the left STG/STS region is not involved in the processing of speech sounds as such, but rather more generally in the processing of the kinds of complex acoustic features that are characteristic to speech sounds.

An important milestone in the research of cortical speech processing was the discovery of "mirror neurons" in the monkey *ventral premotor cortex* (vPMC; di Pellegrino et al., 1992). Such neurons in area F5 of the premotor cortex (PMC) were found to be activated both when the monkey performed hand and mouth actions, as well as during the observation of similar actions performed by other individuals (di Pellegrino et al., 1992; Ferrari et al., 2003). The existence of similar neurons has been demonstrated also in the human brain in a neuronal circuitry that comprises at least Broca's area, the premotor regions, and the primary motor cortex (Fadiga et al., 1995; Hari et al., 1998; Nishitani and Hari, 2000; Buccino et al., 2001; Nishitani and Hari, 2002). These areas form the putative human *mirror neuron system* (MNS) and, along with the properties of mirror neurons in the monkey brain, provide evidence pointing to a close connection between motor actions and perception, thus supporting one of the main claims of Liberman's motor theory of speech perception; that is, perceiving speech is perceiving gestures. Especially the mirror neurons found in BA, which has been demonstrated to act as the speech motor center (Fadiga et al., 2006), provide strong support for the motor perspectives assuming that there is a link between speech production and speech perception.

Recent studies on the neural basis of speech processing

Since the discovery of the MNS, a number of MEG and fMRI studies have demonstrated, further supporting the motor theory of speech perception, that the frontal brain areas involved in the planning and execution of speech gestures (i.e., the posterior part of the left IFG and the vPMC) are activated during auditory, visual and/or auditory-visual speech perception (e.g., Nishitani and Hari, 2002; Wilson et al., 2004; Ojanen et al., 2005; Pekkola et al., 2005; Skipper et al., 2005; Pulvermuller et al., 2006; Wilson and Iacoboni, 2006; Skipper et al., 2007).

In addition, recent single-pulse *transcranial magnetic stimulation* (TMS) studies show that lip or tongue motor-evoked potentials are enhanced during both passive speech listening and viewing, when stimulating the corresponding area of the left primary motor cortex (Sundara et al., 2001; Fadiga et al., 2002; Watkins et al., 2003; Watkins and Paus, 2004; Roy et al., 2008). This increased excitability of the motor system during speech perception is related to an increase in activity in Broca's area and the *ventral premotor cortex* (vPMC), as shown by a recent study combining PET with TMS techiques (Watkins and Paus, 2004).

Importantly, this speech motor 'resonance' mechanism (Fadiga et al., 2002) appears to be articulatory specific, motor facilitation being stronger when the recorded muscle and the presented speech stimulus imply the same articulator (Fadiga et al., 2002; Roy et al., 2008). This specificity of the speech motor resonance mechanism is also suggested by two recent fMRI studies showing similar somatotopic patterns of motor activity in the superior portion of the vPMC during both producing and listening to or viewing lips- and tongue-related phonemes (Pulvermuller et al., 2006; Skipper et al., 2007).

Finally, recent electrocortical mapping and repetitive TMS (rTMS) studies showed that a temporary disruption of the activity in the left posterior IFG disrupts the ability of participants to perform phoneme discrimination (Boatman, 2004; Romero et al., 2006) and rhyme judgements (Gough et al., 2005), and that stimulating the superior portion of the left vPMC impacts the performance of participants in an auditory syllable identification task where syllables are embedded in white noise (Meister et al., 2007) as well as in a phoneme segmentation task under normal listening conditions (Sato et al., 2009).

Recent models of speech perception

By suggesting that speech perception involves a specific mapping from the articulatory gestures of the speaker into the motor plans of the listener, the studies described in the previous section argue against the view that speech perception relies exclusively on the auditory system without any role of the motor system, as postulated in purely auditory approaches of speech perception. However, they do not demonstrate that speech perception is solely mediated by an articulatory code and simply determined through auditory-to-motor feedforward or direct mapping mechanisms, as claimed in the motor theory of speech perception.

In contrary, as a framework to explain the mechanism by which motor system activity is understood by the brain as relevant to speech perception, recent neurobiological models of speech perception (Callan et al., 2004; Wilson and Iacoboni, 2006; Skipper et al., 2007) postutale that perisylvian and extrasylvian language areas participate in speech perception as components of a large scale neural network by means of successive sensory-to-motor feedforward and motor-to-sensory feedback projections. In these models, multisensory inputs interact with feedback or efference copy from the motor system involved in speech production, the role of which is to constrain phonetic interpretation of the incoming sensory information.

Among the most recent and influential of such models is the dual-stream model of speech processing (Hickok and Poeppel, 2000, 2004, 2007) in which two functionally distinct neural networks, known as dorsal and ventral streams, process speech information. The ventral stream, which involves structures in the superior and middle portions of the temporal lobe, is involved in the mapping of the acoustic speech signal into acoustic-phonetic representations, and the dorsal stream, which involves structures in the posterior frontal lobe and the posterior dorsal-most aspect of the temporal lobe and parietal operculum, for translating acoustic speech signals into articulatory representations. Furthermore, the model assumes that the ventral stream is largely bilaterally organized, whereas the dorsal stream is strongly left-hemisphere dominant.

Indeed, Hickok and Poeppel (2007) propose that speech perception tasks (referring here to sublexical tasks, such as syllable discrimination) rely to a greater extent on dorsal stream

circuitry, whereas speech recognition tasks (referring to the set of computations that transform acoustic signals into a representation that makes contact with the mental lexicon) rely more on ventral stream circuitry. It is thus assumed that the ventral stream subserves comprehension of meaningful speech, while the dorsal stream provides a link between speech perception and production. To put it plainly, the ventral stream is responsible for translating sound to meaning and the dorsal stream for translating sound to action. However, the dorsal stream is not considered to be a critical component of speech perception under normal listening conditions in adults, but rather an essential part for speech development and normal speech production. Figure 2.20 shows a schematic diagram of the dual-stream model as well as the approximate anatomical locations of its components.

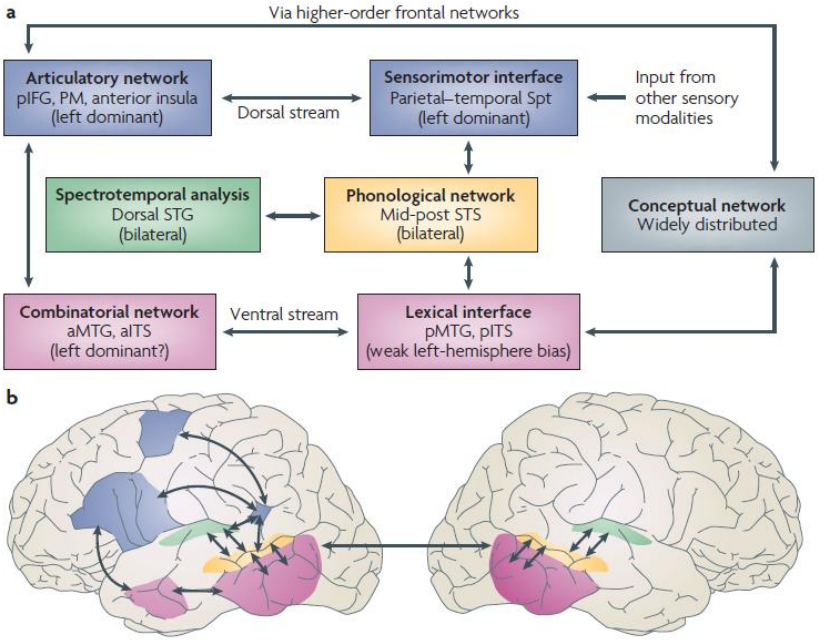

Figure 2.20: a) Schematic diagram of the dual-stream model. Through some form of spectrotemporal analysis (green) and phonological-level processing (yellow), the system diverges in parallel into two streams: a dorsal stream (blue) that maps sensory or phonological representations onto articulatory motor representations, and a ventral stream (violet) that maps sensory or phonological representations onto lexical conceptual representations. b) Approximate anatomical locations of the dual-stream model components. The areas proposed to be involved in spectrotemporal analysis (shaded green) are located in auditory cortices on the dorsal surface of the STG bilaterally. Phonological-level processing areas (shaded yellow) involve the posterior half of the STS bilaterally. The ventral stream (shaded violet) is also bilaterally organized (though with a weak left-hemisphere bias) with the more posterior regions (i.e., posterior middle and inferior portions of the temporal lobes) corresponding to the lexical interface, linking phonological and semantic information, and the more anterior area corresponding to a proposed combinatorial network of speech processing. Regions shaded blue represent the strongly left dominant dorsal stream. The smaller posterior region of the dorsal stream corresponds to an area in the Sylvian fissure at the parietotemporal boundary, which is proposed to be a sensorimotor interface, whereas the larger, more anterior areas in the frontal lobe, involving Broca's region and a more dorsal premotor site, correspond to portions of the articulatory network. Adapted from Hickok and Poeppel (2007).

Another influential model proposed by Skipper and colleagues (2007) claims that early multisensory speech representations in the posterior part of the left STG/STS can be thought of as multisensory hypotheses about the phonemes produced by a speaker. These hypotheses, extracted from sensory information, are translated onto motor control commands (the motor goal of the movement) used in speech production. These commands, localized in the IFG, could then, based on past articulatory experience, generate corresponding motor actions in the vPMC and MC. Activated motor commands would then predict the acoustic and somatosensory consequences of executing a speech movement through efference copy to both the left STG/STS and somatosensory cortices, respectively. Finally, these internally generated sensory consequences are thought to influence or constrain the ultimate phonetic interpretation of the incoming sensory information.

Altogether, the recent models of speech processing argue against the view that speech perception relies exclusively on the auditory system and the acoustic properties of speech but also that speech perception is determined only through feedforward mechanisms from auditory to motor regions. Rather, speech perception is best conceptualized as an interactive neural process involving reciprocal connections between sensory and motor areas whose connection strengths vary as a function of the perceptual task and the external environment (Sato et al., 2009). This is the view defended in the “Perception-for-Action-Control Theory” (PACT) (Schwartz et al., 2002, 2007, in press) in which a speech gesture is not considered as a pure articulatory unit, but rather as a motor coordination shaped by motor-to-sensory nonlinearities.

This theory, centered on the co-structuring of the perception and action systems in relation with phonology, is clearly different from both an auditory theory in which the sensory-interpretative chain is considered independently of the patterning of sounds by speech gestures, in the search of some "direct link" between sounds and phonemes, and from a motor theory in which perception is nothing but a mirror of action, in the claim of a "direct link" between sounds and gestures. It is rather focused on multimodal percepts regularized by motor constraints, in which motor representations should play a crucial role in shaping perceptual units, predicting future sensory events or integrating events in a hopefully smart way, complementing them with adequate articulatory information. Speech perception is therefore thought as the set of perceptual processes allowing to recover and specify the

timing and targets of speech gestures, or, as Wilson and Iacoboni (2006) put it, “neither purely sensory nor motor, but rather a sensorimotor process”.

2.6 Aim of the study

Whereas neuroimaging and neurophysiological data clearly argue for the existence of sensory-to-motor feedforward pojections in speech perception, the evidence for subsequent motor-to-sensory feedback projections, however, is sparse.

The study of this thesis was carried out as a preliminary part of a larger project, in which the aim is to further test the existence of feedback motor-to-auditory projections in auditory speech perception.

The aim of the study of this thesis, however, was to test the hypotheses of the two distinct theoretical perspectives (i.e., auditory and motor) on speech perception, and coincidentally do the groundwork regarding the larger project, by investigating how two different stimulus types (with different levels of ambiguity) presented in four conditions (with varying subsequent-to-stimuli motor tasks required from the subjects) affect the neural mechanisms underlying speech perception. More specifically, by performing a fairly straightforward analysis on the MEG data, the aim was to observe whether the amplitudes, latencies, locations, and directions of the modeled sources of the underlying current distributions change with the stimulus type and/or motor task.

Concerning the auditory / motor theory hypothesis testing, we specifically hypothesized that the perception of the degraded speech stimuli would imply a greater motor involvement and that the neural responses in the conditions with differing subsequent-to-stimuli motor task would show a hierarchy of motor involvement, which could be manifested, for instance, by the modeled center of gravity of the neuronal activity shifting to more posterior position with the more active tasks. These kinds of results would argue against the purely auditory approaches, claiming that speech perception relies exclusively on the auditory system, and, coincidentally, argue for the conception that the speech motor system has a role in speech perception.

Chapter 3

Methods

3.1 Subjects

The test subjects of the MEG experiment included 12 healthy individuals (5 females, 1 left-handed), ranging in age from 21 to 58 (mean ± stdev = 29.5 ± 9.9). All of the subjects were native Finnish-speakers and reported normal hearing ability. The subjects were offered a payment for participating in the experiment. The data of two subjects had to be excluded from the analysis due to technical difficulties, thus making the number of subjects included in the analysis N=10 (3 females, 1 left-handed; aged 24-58, mean ± stdev = 31.0 ± 10.3).

3.2 Stimuli

The stimuli used in the experiment consisted of five individual utterances of both /pa/ and /ta/ syllable sounds, produced by a native male Finnish speaker. The ten selected, clearly articulated tokens were cut 100ms preceding the detected consonantal burst and 100ms following it (at zero crossing points) and finally scaled to 68 db. A silence period was added to the end of each token to obtain a stimulus length of 300ms. In total, five sets of ten individual tokens (five /pa/ and five /ta/) were generated, each with different characteristics (described in the following paragraph).

According to the hypothesis that the speech motor centers are strongly recruited when the mapping between auditory information and phonetic categories is not sufficiently deterministic, masked syllables were created for the last four sets. The mask consisted of Gaussian white-noise that had a 5 ms rise decay envelope and was de-emphasized to better match the frequency spectrum of /pa/ and /ta/ syllables (at -6db/oct). The duration of the mask was 100 ms and was presented simultaneously from the consonantal burst to the beginning of the silence period of the stimuli, with SNRs of +5dB, 0dB, -5dB, and -7dB for sets two through five, respectively. Each noise was cut from a zero crossing point and was

ramped down over 5 ms preceding the silence period. Figure 3.1 shows example waveforms and spectrograms of one individual /pa/ and one individual /ta/ syllables with and without masking. The mask of the figure is with +5 dB SNR.

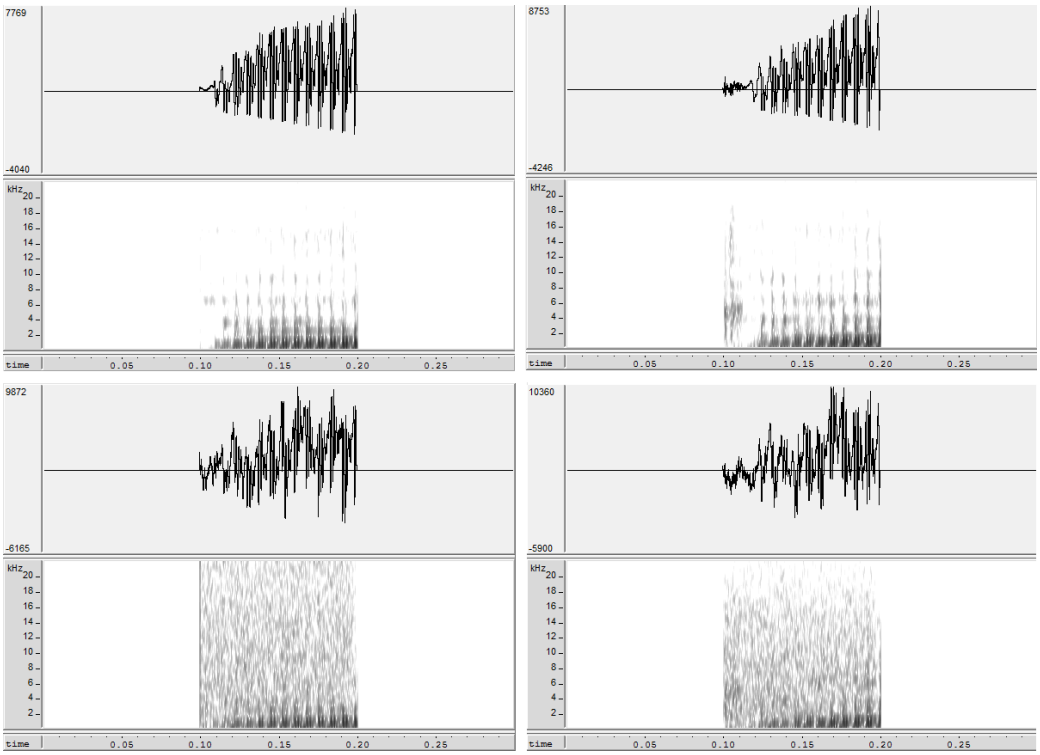

Figure 3.1: Examples of the waveforms and spectrograms of the auditory stimuli used in the study. The left side of the figure depicts an individual /pa/ syllable and the right side depicts an individual /ta/ syllable, without masks (on the top) and with +5dB masks (on the bottom). The figure also illustrates the timing, showing the durations of the syllable sound (and masks) as well as the silence periods.

However, for the experiment, only one of the four sets of stimuli with syllables embedded in noise was to be chosen along with the noiseless set. For this purpose, discrimination tests with a two-forced-choice procedure between /pa/ and /ta/ were conducted and carried out with six healthy Finnish-speaking subjects with self-reported normal hearing ability. Each test set had 50 randomized trials (25 /pa/ and 25 /ta/) and were presented via headphones with a sound level of 65dB. Table 3.1 shows the results.

Further tests were conducted with the +5dB and 0dB SNR sets using 55dB sound level. The results are shown in Table 3.2.

Table 3.1: Percentages of correct responses for each noise set at 65dB sound level.

	SNR +5 dB	SNR 0 dB	SNR -5dB	SNR -7dB
Subject 1	92 %	62 %	58 %	58 %
Subject 2	88 %	52 %	60 %	54 %
Subject 3	86 %	70 %	60 %	50 %
Subject 4	54 %	84 %	76 %	54 %
Subject 5	56 %	60 %	56 %	48 %
Subject 6	86 %	58 %	48 %	44 %
Mean	77 %	64 %	60 %	51 %

Table 3.2: Percentages of correct responses for the +5dB and 0dB noise sets at 55 dB.

	SNR +5dB @ 55 dB	SNR 0dB @ 55dB
Subject 1	98 %	92 %
Subject 2	96 %	92 %
Subject 3	100 %	74 %
Subject 4	60 %	40 %
Subject 5	90 %	70 %
Subject 6	92 %	76 %
Mean	89 %	74 %

Because the noisy syllables for the purpose of the experiment needed to be such that they were on the verge of perception (i.e., not too easy nor too hard to correctly perceive), the decision was made to use the +5dB SNR set at 65dB (with the 77% hit rate).

3.3 Experimental setup

The experiment consisted of four conditions, labeled *passive perception*, *repetition*, *mental repetition*, and *imitation*. Each condition comprised the same stimuli; that is, five individual noiseless /pa/ syllables, five individual noiseless /ta/ syllables, five individual /pa/ syllables embedded in noise with +5dB SNR, and five individual /ta/ syllables embedded in noise with +5dB SNR. In each condition, these stimuli were randomly presented in total of 300 times, so that each individual stimulus was presented 20 times. Because the duration of one trial

(enclosing one stimulus) was six seconds, the conditions were divided into two parts, each with 150 trials and a duration of 15 minutes. To further ensure alertness of the subjects, it was decided to split the tests on two days, so that passive perception and repetition were on one day, and mental repetition and imitation on another.

In addition to the auditory stimulus, a trial included a visual fixation cross, which had a two-fold purpose. First, to keep the eyes of the subjects fixated on one spot, thus preventing irrelevant brain activity when measuring with MEG. Second, to inform the subject when to respond by briefly changing the color of the otherwise black cross to red two seconds from the beginning of a trial (not relevant in the passive perception condition). The duration of the red cross (fixation cue) in a trial was 200 milliseconds. The onset time of the sound stimulus was randomly varied between 1-1.5 seconds from the beginning of a trial to prevent the subject from predicting the stimuli, thus ensuring a more reliable elicitation of the N100m AEF component. The design of a trial is shown in the Figure 3.2. The duration of the trial was stretched after the pilot test from 4 to 6 seconds by adding 2 extra seconds to the duration between the (possible) response of the subject and the beginning of the subsequent trial, thus achieving a longer interstimulus interval (ISI) and, to same extent, a longer neuronal recovery time. The trials were presented with Presentation software (Neurobehavioral Systems Inc., Albany, CA, USA).

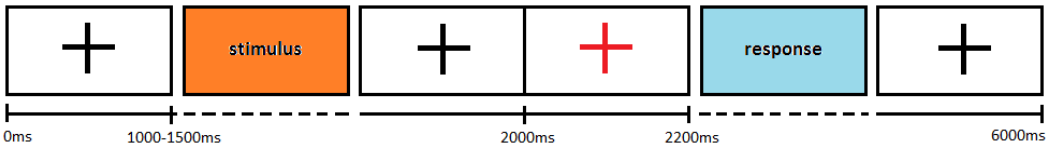

Figure 3.2: The design of a trial used in the experiment.

While in the passive perception condition the subjects were merely to keep their gaze on the fixation cross, the other three conditions required subject response after the visual sign described above. In the repetition condition, the subjects were to simply repeat the syllable they heard; that is, utter either /pa/ or /ta/ accordingly to their auditory perception. The corresponding response in the mental repetition condition for the subjects was to repeat the syllable in their minds; that is, to mentally produce the syllable without any articulatory movements or sound production. In the imitation condition, the subjects were, as the label

suggests, told to imitate the sound stimulus. A notable distinction to the repetition condition is that the subjects were not limited to utter exclusively /pa/ or /ta/, but to imitate exactly what they had heard, even though it might have been something else than /pa/ or /ta/.

As mentioned before, the aim of the study was to test the involvement of the speech motor system in speech perception. According to the underlying hypothesis behind the selection of the aforementioned four experiment conditions, the involvement of the speech motor system was thought to be stronger in the conditions where some sort of successive production after the perceived syllable was involved. More precisely, the involvement in the repetition condition was thought to be stronger than that of in the passive perception condition, yet not as strong as in the imitation condition. The mental repetition condition was established as a sort of control condition, with the assumed similar motor activity with the repetition condition but the possibility of the produced sound of the subject affecting the results ruled out. In other words, the interstimulus intervals in the repetition and imitation conditions are half of those in the passive perception and mental repetition conditions (due to the fact that the subjects hear their own voices repeating [or imitating] the stimuli), making it three and six seconds, respectively.

Subjects were seated in the magnetically shielded room (MSR) on the MEG chair in front of a screen with 30.4” color picture. The distance of the screen was about 170cm from the head of the subject. The fixation cross was presented in the middle of the screen on white background. The size of the cross was 92mm x 92mm, with arm thickness of 7mm. The auditory stimuli were presented via a panel speaker located up by the roof on the opposite wall to the subject at a distance of about 280cm from the head of the subject. The sound level with which the stimuli were presented was about 65dB.

Microphone recordings, in which both the auditory stimulus and the response of the subject are audible, were made in the repetition and mental repetition conditions (apart from the repetition condition in the case of one subject). The recordings were made with Windows Sound Recorder with radio quality (i.e., mono sound with 22050 Hz sample rate and 8-bit sample size).

Subjects were told to keep their gaze fixed on the fixation cross and remain still. Furthermore, they were asked to avoid blinking as much as possible (especially right after

the occurrence of the sound stimulus). A few minute breaks with the possibility to drink water or coffee were held between the 15-minute measuring blocks in order to maintain alertness of the subjects.

3.4 Data acquisition

MEG data were recorded with a 306-channel whole-scalp neuromagnetometer (Neuromag Vectorview, Helsinki Finland) with 102 sensor elements in a helmet array (two orthogonal planar gradiometers and one magnetometer in each element) located at Low Temperature Laboratory, Helsinki University of Technology. The device is situated in a magnetically shielded room, covered with two layers of μ-metal and aluminum to attenuate any effects of outside magnetic fields. The recording and analysis software used in the experiment was provided by the equipment manufacturer, Neuromag Ltd (Espoo, Finland). In addition to all of the 306 MEG channels, one electro-oculographic (EOG), three electromyographic (EMG), and a microphone channel were put to use in the experiment (see the next paragraph).

For artifact detection, a number of electrodes were attached to the face of the subject. First, for detecting eyeblinks as well as vertical and horizontal eye movements, two EOG electrodes were attached: one right below the lower left eyelid and one beside the left eye on the temple, near the outer canthus. To record the position of the head during the exam, four marker coils were placed on the head of the subject with the positions determined in relation to three anatomical landmark points (the nasion and both preauricular points) using an Isotrak 3D-digitizer.

The MEG signals were bandpass filtered at 0.03–200Hz and digitized at a sampling frequency of 2000Hz. In the online-averaging, epochs with over 3000 fT/cm amplitude in the MEG channels or 150 μV in the EOG channel were automatically rejected from the average. The epoch time limits were set to 100ms preceding the stimulus (i.e., a syllable or the red cross) and 600ms following it.

After the acquisition of the online-averaged evoked responses, a further averaging was done by combining the responses for all the noise and no-noise syllables (/pa/s and /ta/s together) of each of the 15-minute measuring blocks into separate sets, and then further

combining the averaged sets of the two 15-minute blocks within each condition, thus resulting in the total of eight averaged response sets per subject, two (no-noise and noise) for each of the four condition, each with up to 150 summed responses (however, due to rejections during the online-averaging the actual number of summed artifact-free responses was generally around 130, with slight variabilities between subjects). The actual computations were done using a weighted average with respect to the number of responses.

3.5 Data analysis

After the averaging, the MEG responses were low-pass filtered at 30 Hz (with 5 Hz as the width of the transition region of the filter) and baseline-corrected 100 ms pre-stimulus. The source localization was done using a dipole model with single ECD fitted in a least-squares sense at the individually determined peak latency of the N100m response, using a fixed subset of 34 planar gradiometers over the left hemisphere temporal areas. A similar, supplementary procedure for comparative purposes was performed also regarding the right hemisphere. A spherical head model was used.

The determination of the N100m peak latencies with the maximum amplitude was done with Elekta Neuromag Source Modelling software (also called Xfit) using a procedure where a single dipole is fitted to the data at evenly spaced intervals. The interval was set to 2ms and the fitting time range beginning and ending 20ms before and after the perceived approximation of the peak latency, respectively. In cases where the ECD had no unambiguous local maximum in source moment, a local maximum of a value called goodness-of-fit was chosen as representative of the source. The average goodness-of-fit values of the ECDs accepted for analysis was 88.7%.

Differences between the latencies, amplitudes, locations, and orientations of the ECD responses were statistically tested using a two-way repeated measures analysis of variance (ANOVA) with condition (passive perception vs. repetition vs. mental repetition vs. imitation) and stimulus type (no-noise vs. noise) as within-subjects factors. The analysis was done with Statistica software (version 9, StatSoft Inc., Tulsa, USA).

Chapter 4

Results

4.1 Dipole source waveforms

Figure 4.1 shows the single ECD source waveforms fitted in both hemispheres. The figure depicts the waveforms of both stimulus types presented in all conditions and averaged over all subjects (N=10). Between the subjects, the peak latencies of the N100m component varied roughly from 90 ms to 140 ms (note that the actual measured latencies were 100ms longer due to the equivalent time period preceding the consonantal burst in the stimuli, see Figures 4.4 and 4.8). The variation of the corresponding dipole moments was from 10 nAm to 110 nAm.

The most apparent difference between the four waveforms is the difference between the two hemispheres. That is, amplitudes of the N100m components in the right hemisphere are about two thirds of those in the left hemisphere. The successive P200m components show a similar effect only in the mental repetition and imitation conditions. A slight over-all difference in the latencies between the hemispheres is also visible.

Further visual inspection of the waveforms reveals differences in the ERFs between the no-noise and noise stimulus types; that is, the latencies of the N100m components with the noise stimuli are shorter and the amplitudes larger than with the no-noise stimuli in all four conditions (and in both hemispheres). In addition to the differences in the N100m components, an inspection of the successive P200m components also show clear differences between the no-noise and noise stimulus types; that is, the amplitudes with the no-noise stimuli in the left hemisphere are distinctly larger in all four conditions (even when measured from the preceding N100m peak). The P200m components of the right hemisphere or the P100m components in both hemispheres, on the other hand, show no such trend. Moreover, differences between the four conditions are substantially harder to identify with visual inspection than those of the stimulus types.

Figure 4.2 shows single-subject left hemisphere ECD fits at N100m peak latency with field maps and an arrow depicting the estimated source strengths and orientations of the dipoles. Along with differences of the N100m component moments between the no-noise and noise stimulus types (with the imitation condition being an exception in this case), visual inspection of the dipole arrows reveals also orientational differences between the two types of stimuli.

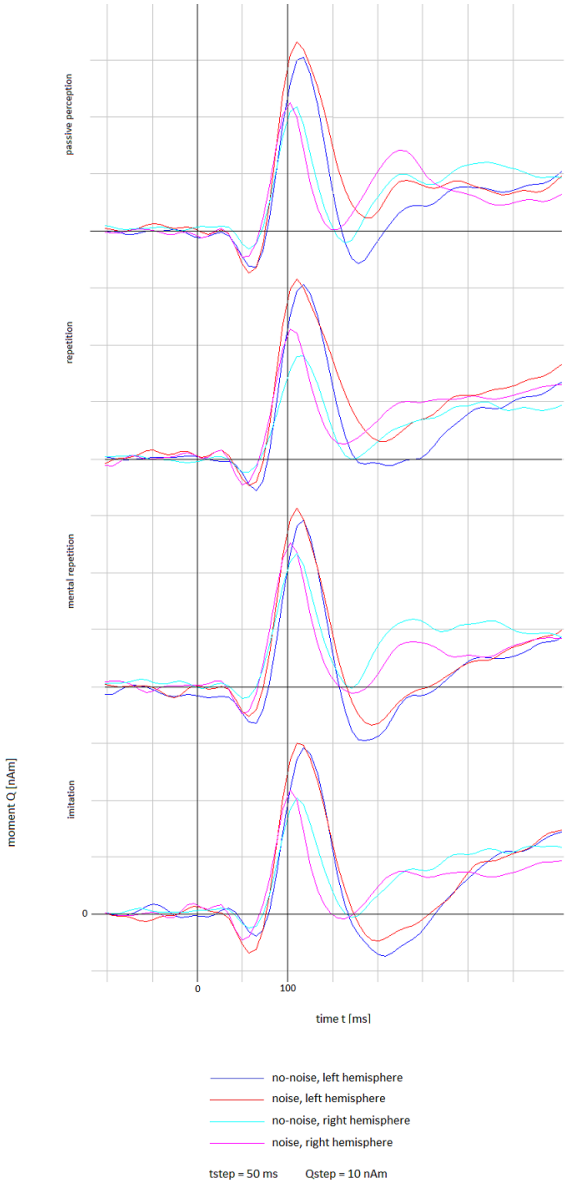

Figure 4.1: The ECD source waveforms averaged over all subjects. A 100ms corrective reduction has been applied to the actual peak latencies on the time axis due to the equivalent time period preceding the consonantal burst in the stimuli.

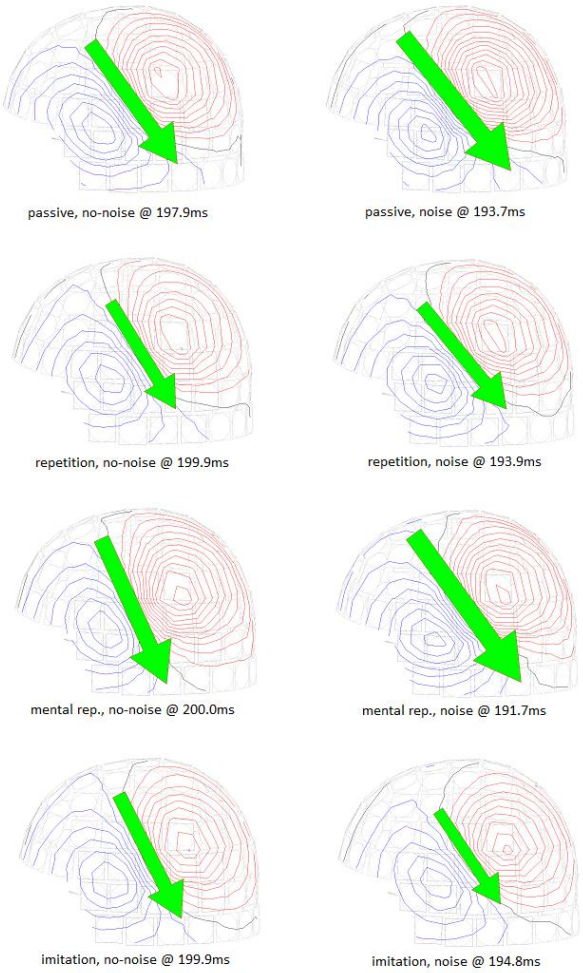

Figure 4.2: Single-subject left-hemisphere ECD fits at N100m peak latency of all four conditions with both stimulus types, the arrow depicting estimated source strengths and orientations. The respective N100m peak latencies are shown below each illustration (note that these measured latencies are 100ms "too long" due to the equivalent time period preceding the consonantal burst in the used stimuli).

4.2 Attributes of the N100m ECDs

From the dipole fitting results, the amplitude (moment), latency, source location, and orientation of the N100m ECD were chosen for inspection, with an emphasis, however, on the location and orientation. Average results (N=10) with the standard error of mean (SEM) are shown in Figures 4.3-4.6, plotting *condition x stimulus type x hemisphere* interactions for the four chosen factors of interest.

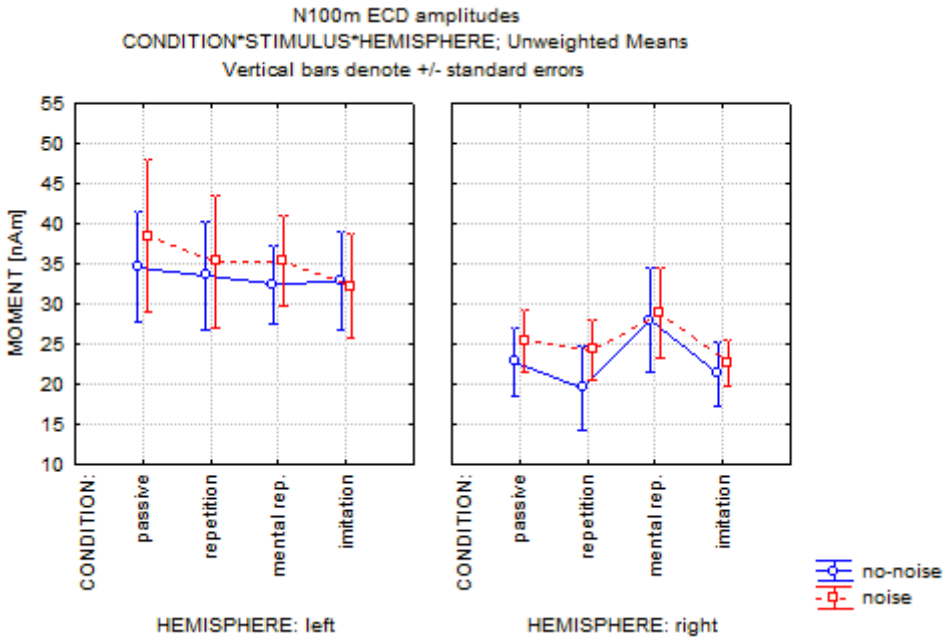

Figure 4.3: Amplitudes (moments) and corresponding SEMs N100m ECDs in both hemispheres with the two stimulus types (no-noise and noise) in the four experimental conditions (see also Figure 4.7).

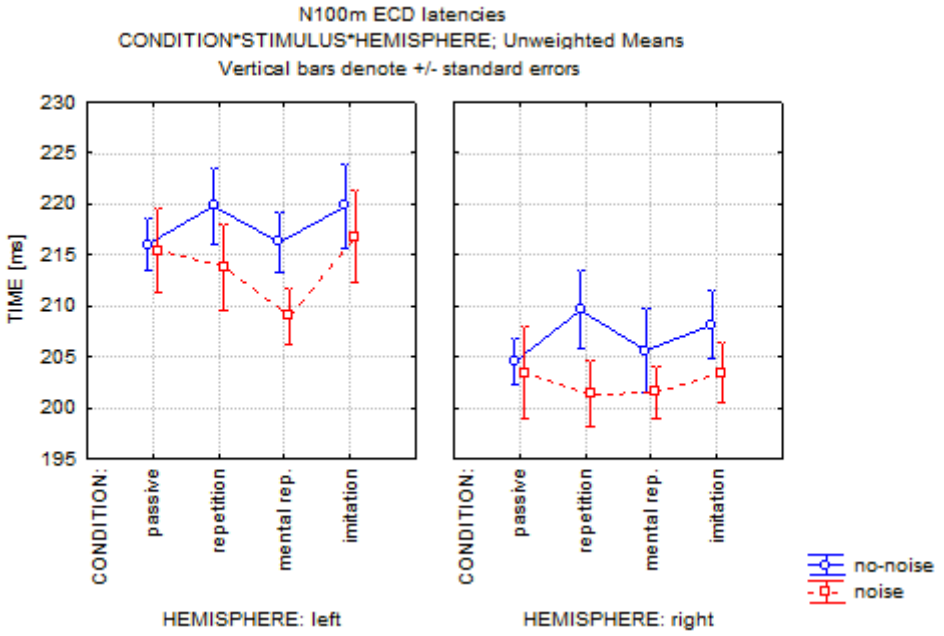

Figure 4.4: Latencies and corresponding SEMs of the N100m ECDs in both hemispheres. The apparent lateness of the N100m peak latencies is due to the 100ms time period preceding the consonantal burst in the stimuli (see also Figure 4.8).

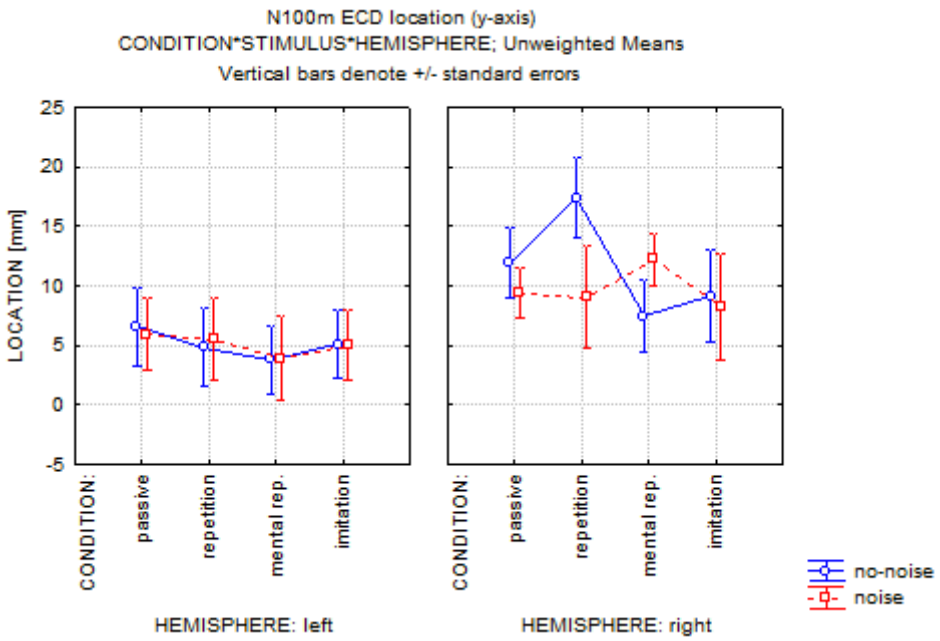

Figure 4.5: Locations and corresponding SEMs the N100m ECDs in both hemispheres along the y-axis the head coordinate system (see Figures 2.6 and 4.9).

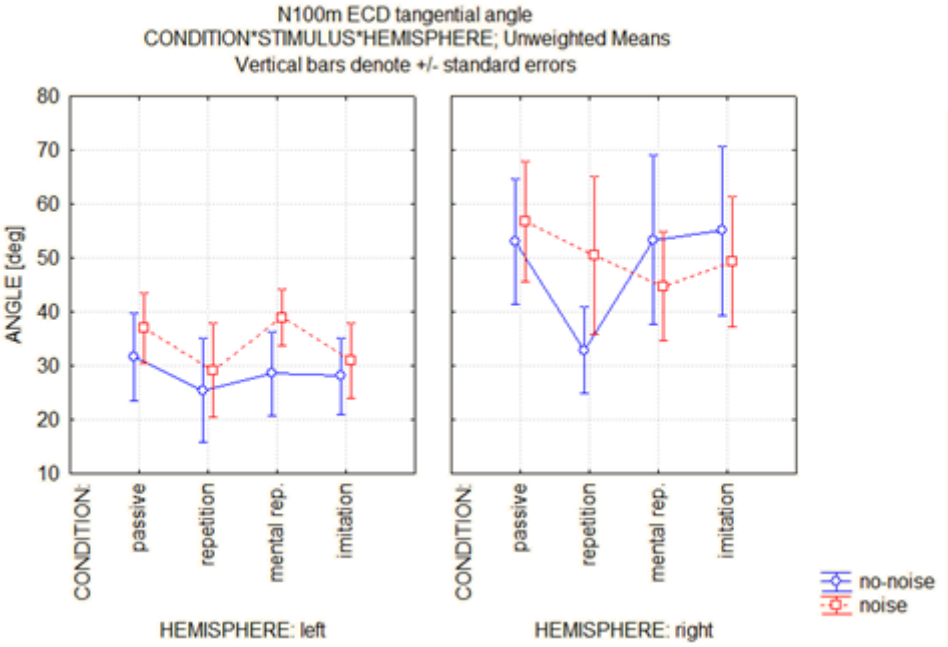

Figure 4.6: Tangential angles and corresponding SEMs of the N100m ECDs in both hemispheres (see also Figure 4.10). The tangential angle depicts the orientational angle between the dipole and the e_θ -unit vector corresponding to the dipole point in sphere model coordinates; that is, if the angle is 0, the dipole points down from the vertex while with 90 degrees, the orientation is around the head in counterclockwise direction (i.e., in the case of the left hemisphere, 90 degrees points backwards in the negative y-axis direction).

For a more thorough and itemized understanding, Figures 4.7-4.10 show the differences between conditions and stimulus types are plotted separately (i.e., *condition x hemisphere* and *stimulus x hemisphere* interactions).

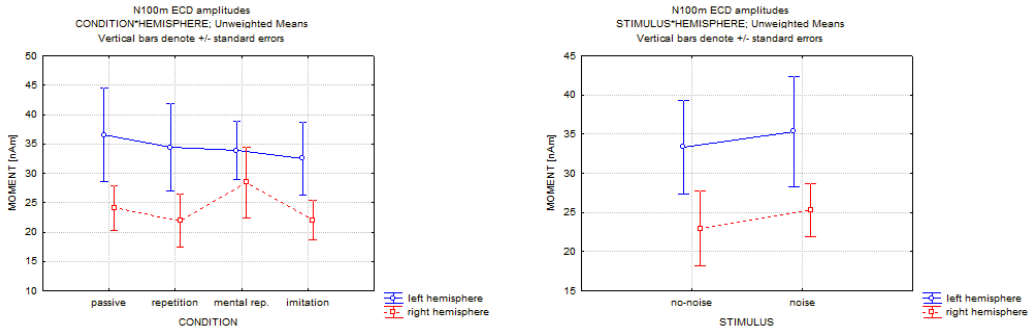

Figure 4.7: Amplitudes of the N100m ECDs in both hemispheres, with the effects for condition and stimulus types plotted separately.

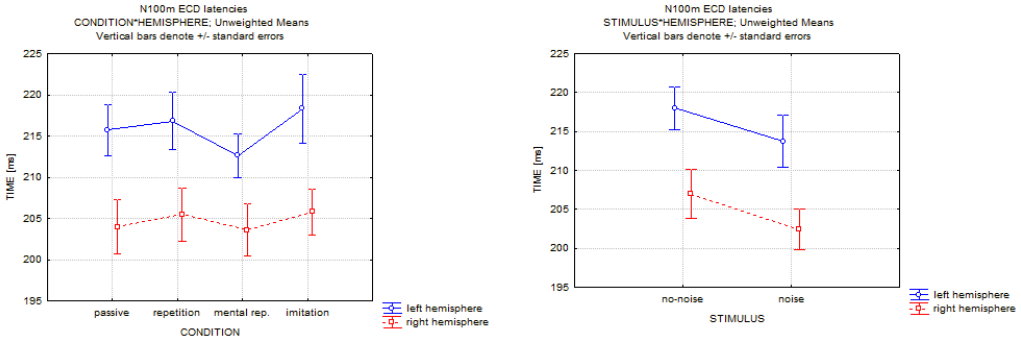

Figure 4.8: Latencies of the N100m ECDs in both hemispheres, with the effects for condition and stimulus types plotted separately.

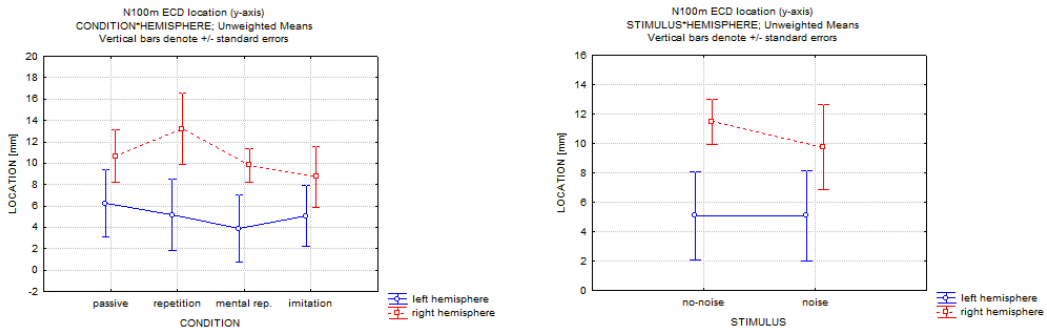

Figure 4.9: Locations of the N100m ECDs in both hemispheres along the y-axis of the head coordinate system (see Figure 2.6), with the effects for condition and stimulus types plotted separately.

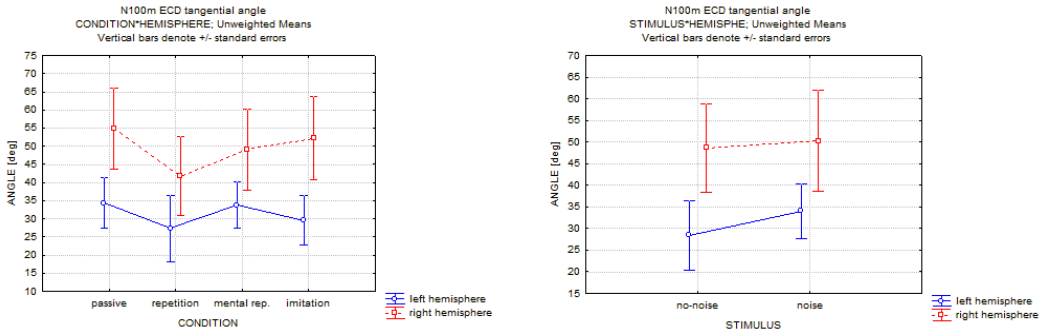

Figure 4.10: Tangential angles of the N100m ECDs in both hemispheres, with the effects for condition and stimulus types plotted separately. 0° angle designates the dipole pointing down from the vertex while with 90°, the orientation is around the head in counterclockwise direction (here, backwards in the negative y-axis direction).

4.3 Statistical analysis

Table 4.1 presents the F-values and corresponding p-levels of the *condition x stimulus type* interactions in both hemispheres. As can be seen in Figures 4.3-4.6, the interactions show some differences and trends either between conditions or stimulus types especially in the left hemisphere, yet as a whole, they fail to reach the level of statistical significance (i.e., $p \leq 0.05$, indicating that there is a 5% probability that the relation between the variables in the tested sample is a "fluke"). However, the interactions for the left hemisphere N100m ECD orientation and the right hemisphere N100m ECD location show near significant effects with $p=0.09$ and $p=0.10$, respectively.

Table 4.2 shows the F-values and corresponding p-levels of the main effects for the four factors. The only main effects to reach the level of statistical significance were the effects of stimulus for the latency of the N100m component in both hemispheres (with $p=0.015$ for both hemispheres).

Table 4.1: F-values and p-levels of the *condition x stimulus type* interactions of the N100m ECD.

Factor	Hemisphere	F(3,27)-value	p-level
amplitude	left	0,45	0,72
	right	1,03	0,40
latency	left	1,23	0,32
	right	1,12	0,36
location	left	0,33	0,81
	right	2,33	0,10
orientation	left	2,44	0,09
	right	1,27	0,30

Table 4.2: F-values and p-levels of the main effects of condition and stimulus type for the amplitude, latency, location, and orientation of the N100m ECD. Red color indicates statistical significance.

Factor	Effect	Hemisphere	F(3,27)-value	p-level
amplitude	condition	left	0,47	0,7
		right	1,69	0,19
	stimulus	left	1,14	0,31
		right	1,86	0,21
latency	condition	left	1,58	0,22
		right	0,47	0,7
	stimulus	left	9,1	0,015
		right	8,97	0,015
location	condition	left	1,42	0,26
		right	0,69	0,57
	stimulus	left	0,00001	0,997
		right	0,3	0,6
orientation	condition	left	1,64	0,2
		right	1,18	0,34
	stimulus	left	3,2	0,11
		right	0,05	0,84

After initial testing, separate *a priori* (planned) comparison tests (contrast analyses) were carried out to test the differences in the four factors both between the two stimulus types in each condition and between the four conditions (i.e., passive vs. repetition, passive vs. mental repetition, passive vs. imitation etc.) with both stimulus types. Tables 4.3 and 4.5 show the left hemisphere results and Tables 4.4 and 4.6 the equivalent right hemisphere results, listing the F(1,9)-value and p-level of each individual test.

With regards to the left hemisphere, the stimulus type comparison tests (Table 4.3) showed statistical significance for the peak latency ($F(1,9)=12.2$; $p=0.007$) as well as for the orientation ($F(1,9)=6.96$; $p=0.027$) of the N100m ECD in the mental repetition condition. Statistical significance for the peak latency was found also in the condition comparisons (Table 4.5) between the passive and mental repetition as well as between the mental repetition and imitation conditions when using noise stimuli (the respective values being $F(1,9)=5.46$; $p=0.04$ and $F(1,9)=6.45$; $p=0.03$). Further findings included a statistical significance for location when comparing the passive and mental repetition conditions with no-noise stimuli ($F(1,9)=5.77$; $p=0.04$), along with a near-significance for orientation when comparing mental repetition and imitation with noise stimuli ($F(1,9)=4.50$; $p=0.06$).

The respective right hemisphere results showed statistical significance for the N100m latency when comparing the no-noise and noise stimulus types in the repetition condition ($F(1,9)=8.07$; $p=0.02$). Notably, the only other near-significant statistical differences (i.e., $p<0.1$) in the stimulus type comparisons were also in the repetition condition for all the other factors (i.e., amplitude, location, and orientation). This effect can be verified in the graphs (on the right side) of Figures 4.3-4.6.

Further statistical significances were found in the right hemispheric condition comparison tests for the ECD orientation between the passive and repetition conditions when using no-noise stimuli ($F(1,9)=6.88$; $p=0.03$) as well as between the passive and mental repetition conditions when using noise stimuli ($F(1,9)=7.42$; $p=0.02$). Statistical significance was found also for amplitude between the repetition and mental repetition condition (highly significant level; $F(1,9)=21.5$; $p=0.001$) and for latency between the passive and imitation conditions ($F(1,9)=7.41$; $p=0.02$), both when using no-noise stimuli. As evident, more statistical significances were found in the right hemisphere results, however, the results are much more incoherent than those of the left hemisphere (as can be perceived in Figures 4.3-4.10), showing no hierarchical effects of motor activity (as expected).

Table 4.3: F-values and p-levels of the left hemisphere comparison tests for testing differences between no-noise and noise –stimulus types in all conditions. Red color indicates statistical significance.

Factor	Condition	F(1,9)-value	p-level
amplitude	passive	0,997982	0,343899
	repetition	0,845774	0,381739
	mental rep.	0,732307	0,414338
	imitation	0,040931	0,844171
latency	passive	0,047162	0,832921
	repetition	2,625653	0,139601
	mental rep.	12,24520	0,006731
	imitation	2,276979	0,165584
location	passive	0,366562	0,559840
	repetition	0,319975	0,585447
	mental rep.	0,005726	0,941337
	imitation	0,005121	0,944515
orientation	passive	1,291369	0,285148
	repetition	3,993453	0,076755
	mental rep.	6,964907	0,026953
	imitation	0,730566	0,414872

Table 4.4: F-values and p-levels of the right hemisphere comparison tests for testing differences between no-noise and noise –stimulus types in all conditions. Red color indicates statistical significance (i.e., $p < 0.05$), blue is for near-significant values with $0.05 \leq p < 0.1$.

Factor	Condition	F(1,9)-value	p-level
amplitude	passive	0,910913	0,364802
	repetition	4,284952	0,068363
	mental rep.	0,157111	0,701064
	imitation	0,633257	0,446643
latency	passive	0,170754	0,689114
	repetition	8,073168	0,019361
	mental rep.	2,094293	0,181767
	imitation	2,440345	0,152686
location	passive	2,580498	0,142650
	repetition	4,184651	0,071113
	mental rep.	1,201678	0,301451
	imitation	0,018545	0,894675
orientation	passive	0,561285	0,472869
	repetition	3,678674	0,087338
	mental rep.	0,354427	0,566285
	imitation	0,130668	0,726081

Table 4.5: F-values and p-levels of the left hemisphere comparison tests for testing differences between the conditions with both stimulus types. Red color indicates statistical significance (i.e., $p < 0.05$), blue is for near-significant values with $0.05 \leq p < 0.1$.

Condition comparison: passive vs. repetition

Factor	Stimulus type	F(1,9)-value	p-level
amplitude	no-noise	0,074059	0,791656
	noise	1,173964	0,306760
latency	no-noise	1,513951	0,249726
	noise	0,280499	0,609189
location	no-noise	2,087587	0,182400
	noise	0,078979	0,785039
orientation	no-noise	2,719479	0,133531
	noise	1,581143	0,240244

Condition comparison: passive vs. mental rep.

Factor	Stimulus type	F(1,9)-value	p-level
amplitude	no-noise	0,408471	0,538671
	noise	0,226025	0,645814
latency	no-noise	0,016316	0,901167
	noise	5,461333	0,044241
location	no-noise	5,772224	0,039732
	noise	1,900892	0,201277
orientation	no-noise	0,975591	0,349098
	noise	0,456780	0,516121

Condition comparison: passive vs. imitation

Factor	Stimulus type	F(1,9)-value	p-level
amplitude	no-noise	0,272676	0,614150
	noise	1,613172	0,235899
latency	no-noise	2,820357	0,127381
	noise	0,139724	0,717209
location	no-noise	1,733139	0,220548
	noise	0,250456	0,628763
orientation	no-noise	1,180576	0,305481
	noise	2,394433	0,156173

Condition comparison: repetition vs. mental rep.

Factor	Stimulus type	F(1,9)-value	p-level
amplitude	no-noise	0,098989	0,760218
	noise	0,000055	0,994225
latency	no-noise	0,855949	0,379014
	noise	2,301990	0,163519
location	no-noise	0,639744	0,444401
	noise	0,732127	0,414393
orientation	no-noise	0,666641	0,435303
	noise	2,788012	0,129312

Condition comparison: repetition vs. imitation

Factor	Stimulus type	F(1,9)-value	p-level
amplitude	no-noise	0,212831	0,655504
	noise	1,106657	0,320234
latency	no-noise	0,000000	0,999999
	noise	0,378874	0,553451
location	no-noise	0,152939	0,704839
	noise	0,102755	0,755866
orientation	no-noise	0,778754	0,400478
	noise	0,173318	0,686933

Condition comparison: mental rep. vs. imitation

Factor	Stimulus type	F(1,9)-value	p-level
amplitude	no-noise	0,024116	0,880017
	noise	0,500411	0,497227
latency	no-noise	2,066290	0,184431
	noise	6,452035	0,031706
location	no-noise	2,526937	0,146380
	noise	0,489658	0,501764
orientation	no-noise	0,034202	0,857379
	noise	4,504031	0,062805

Table 4.6: F-values and p-levels of the right hemisphere comparison tests for testing differences between the conditions with both stimulus types. Red color indicates statistical significance (i.e., $p < 0.05$), blue is for near-significant values with $0.05 \leq p < 0.1$.

Condition comparison: passive vs. repetition

Factor	Stimulus type	F(1,9)-value	p-level
amplitude	no-noise	1,233752	0,295471
	noise	0,079683	0,784111
latency	no-noise	3,277378	0,103677
	noise	0,239847	0,636032
location	no-noise	3,036992	0,115361
	noise	0,009492	0,924523
orientation	no-noise	6,878497	0,027689
	noise	0,873525	0,374378

Condition comparison: passive vs. mental rep.

Factor	Stimulus type	F(1,9)-value	p-level
amplitude	no-noise	2,402911	0,155521
	noise	0,327938	0,580898
latency	no-noise	0,175058	0,685462
	noise	0,169633	0,690074
location	no-noise	0,700683	0,424223
	noise	3,214371	0,106587
orientation	no-noise	0,000211	0,988725
	noise	7,419226	0,023460

Condition comparison: passive vs. imitation

Factor	Stimulus type	F(1,9)-value	p-level
amplitude	no-noise	0,685147	0,429221
	noise	1,089167	0,323877
latency	no-noise	7,408452	0,023537
	noise	0,000020	0,996486
location	no-noise	0,183728	0,678265
	noise	0,069385	0,798166
orientation	no-noise	0,015352	0,904115
	noise	1,195788	0,302568

Condition comparison: repetition vs. mental rep.

Factor	Stimulus type	F(1,9)-value	p-level
amplitude	no-noise	21,50637	0,001224
	noise	1,926451	0,198542
latency	no-noise	3,878474	0,080422
	noise	0,009554	0,924279
location	no-noise	3,396994	0,098426
	noise	0,915074	0,363760
orientation	no-noise	1,603925	0,237142
	noise	0,897859	0,368103

Condition comparison: repetition vs. imitation

Factor	Stimulus type	F(1,9)-value	p-level
amplitude	no-noise	0,320197	0,585319
	noise	0,362499	0,561981
latency	no-noise	0,269969	0,615888
	noise	4,439053	0,064392
location	no-noise	2,346940	0,159892
	noise	0,365722	0,560281
orientation	no-noise	1,774693	0,215549
	noise	0,110991	0,746649

Condition comparison: mental rep. vs. imitation

Factor	Stimulus type	F(1,9)-value	p-level
amplitude	no-noise	2,945582	0,120246
	noise	1,822960	0,209934
latency	no-noise	0,998812	0,343709
	noise	0,485937	0,503352
location	no-noise	0,509481	0,493457
	noise	1,057311	0,330670
orientation	no-noise	0,214872	0,653981
	noise	0,568943	0,469954

Chapter 5

Discussion

The present study investigated the cortical processing of speech sounds in four conditions with differing motor tasks required from the subjects subsequent to the stimuli. In the conditions, two different speech stimulus types with different levels of signal quality were presented to the subjects; that is, one set of stimuli of spoken /pa/ and /ta/ syllables without noise, and one set of the same syllables with Gaussian white-noise masks that were de-emphasized to better match the frequency spectrum of the syllables. Furthermore, instead of using continuous noise to distort the speech stimuli (as in the majority of earlier studies), transient masking was used; that is, the distortion was only present simultaneously with the stimulus.

As a foreword for this chapter, it has to be stated that the single ECD estimation method used in modeling the neuronal source activation in the present study is at best a crude approximation. Thus, regardless of any lack of statistical significance in the present analysis, more refined analyses might better capture the effects of interest.

5.1 Neural response changes with degradation of speech sounds

Although the levels of statistical significance were only partly reached, the results show an increase in amplitude with a concurrent latency reduction to the degradation of speech sounds (see right sides of Figures 4.7 and 4.8 as well as Table 4.2). This amplification of the AEFs indicates that that the auditory cortices of both hemispheres are highly sensitive to distortion. Both the amplitudes and latencies of the N100m response exhibited hemispheric asymmetry in that the response was significantly stronger and delayed in the left hemisphere (see Figures 4.7 and 4.8). However, the right hemisphere was slightly more sensitive to the degradation of the speech stimuli (see Figure 4.7 and Table 4.2). The

observed hemispheric asymmetry of the response amplitudes is to some extent in accord with a recent study, indicating that unmasked speech stimuli (syllables /pi/ and /ti/) elicited stronger N100m responses in the left hemisphere (Davis et al., 2008). The hemispheric ratio was about the same in both studies, right hemispheric responses being about two thirds of those in the left hemisphere. However, unlike in the present study, the effect in Davis et al. (2008) was not the same for masked stimuli; that is, the speech-shaped white noise employed in the study elicited slightly smaller N100m responses in the left hemisphere. A notable curiosity considering the effects on latencies with degradation of the stimuli is that whereas all the active perception tasks showed significant differences between the masked and unmasked stimuli, in the passive perception condition, the differences were practically non-existing (see Figure 4.4).

The observed amplification effect is in accordance with a recent study, in which the degradation effect was achieved by directly manipulating the acoustic structure of the speech signal (Miettinen et al., 2010). In contrast, previous observations using continuous noise to distort speech sounds resulted in decreased and delayed AEP/AEF responses (Shtyrov et al., 1998; Whiting et al., 1998; Muller-Gass et al., 2001; Martin and Stapells, 2005; Kozou et al., 2005; Kaplan-Neeman et al., 2006). However, unlike in the present study, the amplitude of the N100m response in Miettinen et al. (2010) was significantly stronger in the right hemisphere.

Importantly, Miettinen et al. (2010) further demonstrated that this amplifying effect of stimulus distortion is not related to processing of acoustic features of speech in particular, since it was also observed with spectrally simpler, non-speech stimuli (with the amplifying effect being even clearer). Thus, it seems that the increased activation is reflecting the processing of auditory stimulus features common to both speech and non-speech stimuli. Furthermore, as discussed in the section 2.3.1, the amplitude (and latency) of the N100m component is strongly dependent upon the stimulus characteristics, not on the internal cognitive processes, implying that conclusions on any motor involvement in the masked syllable perception processing should be made with reservation.

Regardless, the present findings on the neural response changes with degradation of speech sounds raise interesting questions for future research. Indeed, given that both the masking methods (i.e., transient and continuous) make the speech stimuli less intelligible,

why is it that the neural response changes are converse to each other? In light of the present study and the study conducted by Miettinen et al. (2010) where transient masking was used, it would seem plausible that the AEP/AEF decrements observed in the above-mentioned studies arise from the use of continuous masking itself. This view is supported by findings indicating that continuous noise causes neuronal adaptation in the auditory cortex, which leads to attenuated and delayed AEP/AEF responses for the masked stimuli; that is, to elicit N100(m) response amplitudes and latencies equal to those measured in the case of unmasked stimuli, the intensity of the stimuli presented over the continuous masking noise has to be elevated (Morita et al., 2006; Billings et al., 2009).

Although the stimuli masking methods used in the present study and in Miettinen et al. (2010) were both transient, a notable difference is that in Miettinen et al. (2010) the acoustic structure of the speech signal was directly manipulated in order to introduce spectrally correlated, signal-dependent distortion to the signal, while in the present study the mask was additive, albeit de-emphasized to better match the frequency spectrum of the syllables. Considering this important distinction (as well as the discussed contradictory neural response changes), the future experiments aiming to obtain a more comprehensive picture of the neural mechanisms related to the perception of degraded speech should incorporate an investigation of the effects of spectrally correlated distortion and uncorrelated additive noise using both continuous and transient masking on AEF responses.

5.2 Neural response changes with differing motor tasks

The results show a slight (not statistically significant) hierarchical decrement of the left hemisphere N100m amplitudes with the more active subsequent-to-stimuli perception tasks (when the responses to both stimulus types are averaged together). That is, conversely to the rate of the assumed motor involvement (or sensorimotor integration processes), the amplitude decreased hierarchically from passive to active perception tasks with the smallest amplitude in the case of subsequent imitation (see Figure 4.7). The corresponding latencies show a concurrent hierarchical delay in the cases of passive perception, overt repetition, and imitation, yet not in the case of mental (covert) repetition (see Figure 4.8).

In the cases of the subsequent repetition and imitation, the decreased amplitude could be explained by the subjects' own speech, which has been shown to reduce reactivity of the human auditory cortex (Numminen et al., 1999). Even so, as the subsequent mental repetition, where the subjects were to silently repeat the presented syllable in their minds, also showed reduction with respect to passive perception, one could argue that the explanation for the hierarchical decrement has to be searched for elsewhere. However, in a related study Numminen and Curio (1999) showed that voiceless 'inner' speaking is also sufficient to reduce the reactivity of the human auditory cortex. Thus, instead of being related to the hierarchy of motor involvement, it seems plausible that the dampened amplitudes of the more active perception tasks can be explained by the subjects' own overt and covert speech modifying utterance-specific processing in the human auditory cortex.

5.3 ECD location and orientation shifts with degradation of speech sounds / differing motor tasks

The results show that the ECD shifts significantly to more posterior position with the more active perception tasks in the left hemisphere (see Figure 4.9). The clearest shift, which is also the only statistically significant one, is between the passive perception and mental repetition conditions with the no-noise stimuli (see Table 4.5). As discussed already in the section 3.3, the mental repetition is probably the most reliable motor task for comparison since it shares the same ISI (6 s) with the passive perception condition. The other two active perception conditions, namely repetition and imitation, have an ISI of only around half of that (i.e., 3-4 s, depending on the timing of the subject's response; see Figure 3.2) due to the fact that they hear their own voice repeating (or imitating) the stimuli. Taking into account all the factors influencing the reliability of the results (i.e., the results of the statistical analysis, N100m dependency on the stimulus characteristics, and the respective ISIs of the four experimental conditions), it is justified to suggest that this is the most significant finding of the present study

A possible explanation for the observation is that the underlying current distribution changes so that there are posterior auditory cortex areas that are activated when motor involvement kicks in during the active perception tasks. Moreover, it could be argued that

this finding is in accord with the hypotheses of the dual-stream model of speech processing (Hickok and Poeppel, 2000, 2004, 2007; discussed in section 2.5.2), which suggests that two functionally distinct neural streams (ventral and dorsal stream) account for perceiving speech. The model holds that the strongly left dominant dorsal stream, which first proceeds from the auditory cortex on the dorsal surface of the STG to an area in the Sylvian fissure at the parietotemporal boundary (see Figure 2.17), provides a link between speech production and perception. Indeed, this small posterior region at the parietemporal boundary is proposed to be a sensorimotor interface, facilitating the translation of acoustic speech signals into articulatory representations. Thus, one could argue that the observed shift of the ECD to a more posterior position in the left hemisphere auditory cortex (along the dorsal stream) can be explained by the dorsal stream being more involved during the conditions with a subsequent motor task than during passive perception.

In addition to the location shift, a statistically significant shift in the ECD orientation was found between the unmasked and masked stimuli in the mental repetition condition (see Table 4.3). Similar, yet not as strong, effect was found also in the other conditions (see Figures 4.6 and 4.10). Since the shift is to a more posterior orientation with degradation of the speech stimuli, this effect can be seen to concur with and support the explanations proposed in the previous paragraph (i.e., a change in the underlying current distribution caused by the stress shifting more on the dorsal stream with the degraded speech stimuli) and, consequently, support the hypothesis that the speech motor centers are strongly recruited when the mapping between auditory information and phonetic categories is not sufficiently deterministic.

Supposing that the above interpretations are correct, these two findings, for their part, argue against the purely auditory approaches to speech perception by suggesting that the speech motor system has a role in speech perception, coincidentally giving support to the motor perspective on speech perception. All in all, the findings of the present study raise interesting questions as well as provide a promising basis for future research on speech perception.

Bibliography

Ackermann H, Wildgruber D, Daum I, Grodd W. (1998). Does the cerebellum contribute to cognitive aspects of speech production? A functional magnetic resonance imaging (fMRI) study in humans. *Neuroscience Letters*, 247:187-90.

Alexander MP, Benson DF, Stuss DT. (1989). Frontal lobes and language. *Brain and Language*, 37:656-691.

Baillet S, Garnero L, Marin G, Hugonin JP. (1999). Combined MEG and EEG source imaging by minimization of mutual information. *IEEE Transactions on Biomedical Engineering*, 46:522-534.

Baillet S, Mosher C, Leahy M. (2001). Electromagnetic brain mapping. *IEEE Signal processing magazine*, 18:14-30.

Bear MF, Connors BW, Paradiso MA. (2001). Neuroscience: Exploring the Brain. LippincottWilliams & Wilkins, 2nd edition.

Billings CJ, Tremblay KL, Stecker GC, Tolin WM. (2009). Human evoked cortical activity to signal-to-noise ratio and absolute signal level. *Hearing Research*, 254:15-24.

Binder JR, Frost JA, Hammeke TA, Bellgowan PS, Springer JA, Kaufman JN, Possing ET. (2000). Human temporal lobe activation by speech and nonspeech sounds. *Cerebral Cortex*, 10:512-528.

Binder JR., Liebenthal E, Possing ET, Medler DA, DouglasWard B. (2004). Neural correlates of sensory and decision processes in auditory object identification. *Nature Neuroscience*, 7:295–301.

Bloom W, Fawcett DW. (1975). *A Textbook of Histology*. Philadelphia: WB Saunders, 10th edition.

Boatman DF. (2004). Cortical bases of speech perception: Evidence from functional lesion studies. *Cognition*, 92:47–65.

Brenner D, Lipton J, Kaufman L, Williamson SJ. (1978). Somatically evoked magnetic fields of the human brain. *Science*, 199:81-83.

Brenner D, Williamson SJ, Kaufman L. (1975). Visually evoked magnetic fields of the human brain. *Science*, 190:480-482.

Broca P. (1861a). Nouvelle observation d'aphémie produite par une lésion de la troisiéme circonvolution frontale [New observations of aphemia produced by a lesion of the third frontal convolution]. *Bulletins de la Société d'Anatomie*, 2(6): 398-407.

Broca P. (1861b). Remarques sur le siége de la faculté du langage articulé, suivies d'une observation d'aphémie [Remarks on the location of the faculty of articulate language, followed by an observation of aphemia (loss of speech)]. *Bulletins de la Société d'Anatomie*, 2(6): 330-357.

Buccino G, Binkofski F, Fink GR, Fadiga L, Fogassi L, Gallese V, Seitz RJ, Zilles K, Rizzolatti G, Freund HJ. (2001). Action observation activates premotor and parietal areas in a somatotopic manner: an fMRI study. *European Journal of Neuroscience*, 13:400-404.

Callan DE, Jones J, Callan A, Akahane-Yamada R. (2004). Phonetic perceptual identification by native- and second-language speakers differentially activates brain regions involved with acoustic phonetic processing and those involved with articulatory–auditory/orosensory internal models. *Neuroimage*, 22:1182–1194.

Carlson NA. (1992). *Foundations of Physiological Psychology*. p. 36. Needham Heights, Massachusetts: Simon & Schuster.

Celesia GG, Brigell MG. (1999). Auditory evoked potentials. In: Niedermeyer E, Lopes da Silva F (editors), *Electroencephalography: Basic Principles, Clinical Applications, and Related Fields*. ch. 55. Lippincott Williams & Wilkins, 4th edition.

Clemente CD. (2007). *Anatomy: a regional atlas of the human body*. p. 499. Lippincott Williams & Wilkins, 5th edition.

Curio G, Neuloh G, Numminen J, Jousmäki V, Hari R. (2000). Speaking modifies voice-evoked activity in the human auditory cortex. *Human Brain Mapping*, 9(4):183-191.

David O, Garnero L, Cosmelli D, Varela FJ. (2002). Estimation of neural dynamics from MEG/EEG cortical current density maps: application to the reconstruction of large-scale cortical synchrony. *IEEE Transactions on Biomedical Engineering*, 49:975-987.

Davis C, Kislyuk D, Kim J, Sams M. (2008). The effect of viewing speech on auditory speech processing is different in the left and right hemispheres. *Brain Research*, 1242:151-161.

Davis H, Mast T, Yoshie N, Zerlin S. (1966). The slow response of the human cortex to auditory stimuli: recovery process. *Electroencephalography and Clinical Neurophysiology*, 21(2):105-113.

Delattre PC, Liberman AM, Cooper FS, Gerstman LJ. (1952). An experimental study of the acoustic determinants of vowel color. *Word*, 8:195–210.

Diehl RL, Kluender KR. (1989). On the Objects of Speech Perception. *Ecological psychology*, 1:121-144.

Diehl RL, Lotto AJ, Holt LL. (2004). Speech perception. *Annual Review of Psychology*, 55:149-179.

di Pellegrino, G., Fadiga, L., Fogassi, L., Gallese, V., & Rizzolatti, G. (1992). Understanding motor events: a neurophysiological study. *Experimental Brain Research*, 91(1):176-180.

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

Dronkers NF, Baldo JV. (2001). Neural basis of speech production. *International Encyclopedia of the Social and Behavioral Sciences*, pp. 14875-14879.

Dronkers NF, Plaisant O, Iba-Zizen MT, Cabanis EA. (2007). Paul Broca's Historic Cases: High Resolution MR Imaging of the Brains of Leborgne and Lelong. *Brain*, 130:1432–1441.

Dronkers NF, Redfern BB, Knight RT. (2000). The neural architecture of language disorders. In: Gazzaniga MS (editor), *The New Cognitive Neurosciences*. MIT Press, Cambridge, MA.

Dubin M. (2009). Brodmann areas. Retrieved April 19, 2010, from http://spot.colorado.edu/~dubin/talks/brodmann/brodmann.html

Dubuc B. (2010). The brain from top to bottom. Retrieved May 19, 2010, from http://thebrain.mcgill.ca

Duffy JR. (1995). Motor Speech Disorders: Substrates, Differential Diagnosis and Management. C. V. Mosby-Year Book, St Louis, MO.

Fabbro F, Clarici A, Bava A. (1996). Effects of left basal ganglia lesions on language production. *Perceptual and Motor Skills*, 82:1291-1298.

Fadiga L, Craighero L, Buccino G, Rizzolatti G. (2002). Speech listening specifically modulates the excitability of tongue muscles: a TMS study. *European Journal of Neuroscience*, 15:399-402.

Fadiga L, Craighero L, Fabbri-Destro M, Finos L, Cotilon-Williams N, Smith AT, Castiello U. (2006). Language in Shadow. *Social Neuroscience* , 1(2):77–89.

Fadiga L, Fogassi L, Pavesi G, Rizzolatti G. (1995). Motor facilitation during action observation: a magnetic stimulation study. *Journal of Neurophysiology*, 73(6):2608-2611.

Ferrari PF, Gallese V, Rizzolatti G, Fogassi L. (2003). Mirror neurons responding to the observation of ingestive and communicative mouth actions in monkey ventral premotor cortex. *European Journal of Neuroscience*, 17:1703-1714.

Fowler CA. (1986). An event approach to the study of speech perception direct-realist perspective. *Journal of Phonetics*, 14:3-28.

Fowler CA. (1996). Listeners do hear sounds, not tongues. *Journal of the Acoustical Society of America*, 99:1730–1741.

Galantucci B, Fowler CA, Turvey MT. (2006). The motor theory of speech perception reviewed. *Psychonomic Bulletin & Review*, 13(3):361–377.

Ganslandt O, Ulbricht D, Kober H, Vieth J, Strauss C, Fahlbusch R. (1996). SEF-MEG localization of somatosensory cortex as a method for presurgical assessment of functional brain area. *Electroencephalography and clinical neurophysiology. Supplement*, 46:209-213.

Gavit L, Baillet S, Mangin JF, Pescatore J, Garnero L. (2001). A multiresolution framework to MEG/EEG source imaging. *IEEE Transactions on Biomedical Engineering*, 48:1080-1087.

Geschwind N. (1970). The organization of language and the brain. *Science*, 170:940-944.

Godey B, Schwartz D, de Graaf JB, Chauvel P, Liegeois-Chauvel C. (2001). Neuromagnetic source localization of auditory evoked fields and intracerebral evoked potentials: a comparison of data in the same patients. *Clinical Neurophysiology*, 112: 1850–1859.

Goldstein EB. (2007). *Sensation and Perception*. Wadsworth-Thomson Learning, 7th edition.

Gough PM, Nobre AC, Devlin JT. (2005). Dissociating linguistic processes in the left inferior frontal cortex with transcranial magnetic stimulation. *Journal of Neuroscience*, 25:8010–8016.

Grasman RP, Huizenga HM, Waldorp LJ, Bocker KB, Molenaar PC. (2004). Frequency domain simultaneous source and source coherence estimation with an application to MEG. *IEEE Transactions on Biomedical Engineering*, 51:45-55.

Gross J, Ioannides AA. (1999). Linear transformations of data space in MEG. *Physics in Medicine and Biology*, 44:2081-2097.

Hari R. (1999). Magnetoencephalography as a tool of clinical neurophysiology. In: Niedermeyer E, Lopes da Silva F (editors). *Electroencephalography: Basic Principles, Clinical Applications, and Related Fields*. pp. 1107-1134. Baltimore: Williams & Wilkins.

Hari R, Aittoniemi K, Järvinen ML, Katila T, Varpula T. (1980). Auditory evoked transient and sustained magnetic fields of the human brain. Localization of neural generators. *Experimental Brain Research*, 40:237-240.

Hari R, Forss N, Avikainen S, Kirveskari E, Salenius S, Rizzolatti G. (1998). Activation of human primary motor cortex during action observation: a neuromagnetic study. *Proceedings of the National Academy of Sciences of the United States of America*, 95:15061-15065.

Hickok G, Poeppel D. (2000). Towards a functional neuroanatomy of speech perception. *Trends in Cognitive Science*, 4(4):131-138.

Hickok G, Poeppel D. (2004). Dorsal and ventral streams: A framework for understanding aspects of the functional anatomy of language. *Cognition*, 92:67-99.

Hickok G, Poeppel D. (2007). The cortical organization of speech processing. *Nature Reviews Neuroscience*, 8:393-402.

Hämäläinen M, Hari R, Ilmoniemi RJ, Knuutila J, Lounasmaa OV. (1993). Magnetoencephalography — theory, instrumentation, and applications to noninvasive studies of the working human brain. *Reviews of Modern Physics*, 65:413-497.

Hämäläinen MS, Ilmoniemi RJ. (1984). *Interpreting measured magnetic fields of the brain: Estimates of current distributions*. Helsinki University of Technology, Finland. Report nr TKK-F-A559.

Indefrey P, Levelt WJ. (2000). The neural correlates of language production. In: Gazzaniga MS (editor), *The New Cognitive Neurosciences*. MIT Press, Cambridge, MA.

Iversen LL. (1979). The chemistry of the brain. *Scientific American*, 241:134-149.

Ivry RB, Fiez JA. (2000). Cerebellar contributions to cognition and imagery. In: Gazzaniga MS (editor), *The New Cognitive Neurosciences*. MIT Press: Cambridge, MA.

Ivry RB, Keele SW. (1989). Timing functions of the cerebellum. *Journal of Cognitive Neuroscience*, 1:136-152.

Jensen O, Vanni S. (2002). A new method to identify multiple sources of oscillatory activity from magnetoencephalographic data. *Neuroimage*, 15:568-574.

Jerbi K, Baillet S, Mosher JC, Nolte G, Garnero L, Leahy RM. (2004). Localization of realistic cortical activity in MEG using current multipoles. *Neuroimage*, 22:779-793.

Jones J, Callan DE. (2003). Brain activity during audiovisual speech perception: An fMRI study of the McGurk effect. *Neuroreport*, 14(8):1129–1133.

Kanno A, Nakasato N, Nagamatsu K, Iwasaki M, Hatanaka K, Murayama N and Yoshimoto T. (2001). Comparison of source localization for the P30m, P50m, and N100m peaks of the auditory evoked fields. In: Nenonen J, Ilmoniemi RJ, Katila T (Editors), Biomag2000, Proc 12th Int Conf on Biomagnetism , Helsinki Univ of Technology, Espoo, Finland, pp 25-28.

Kaplan-Neeman R, Kishon-Rabin L, Henkin Y, Muchnik C. (2006). Identification of syllables in noise: Electrophysiological and behavioral correlates. *Journal of the Acoustical Society of America*, 120:926-933.

Kaufman L, Okada Y, Brenner D, Williamson SJ. (1981). On the relation between somatic evoked potentials and fields. *International Journal of Neuroscience*, 15:223-239.

Keidel WD, Spreng M. (1965). Neurophysiological evidence for the stevens power function in man. *Journal of the Acoustical Society of America*, 38:191-195.

Kozou H, Kujala T, Shtyrov Y, Toppila E, Starck J, Alku P, Näätänen R. (2005). The effect of different noise types on the speech and non-speech elicited mismatch negativity. *Hearing Research*, 199:31-39.

Kudo N, Nakagome K, Kasai K, Araki T, Fukuda M, Kato N, Iwanami A. (2004). Effects of corollary discharge on event-related potentials during selective attention task in healthy men and women. *Neuroscience Research*. 48(1):59-64.

Kuhl PK. (1993). Innate predispositions and the effects of experience in speech perception: the native language magnet theory. In: de Boysson-Bardies B, de Schonen S, Jusczyk P, MacNeilage P, Morton J (editors). *Developmental Neurocognition: Speech and Face Processing in the First Year of Life*. pp. 259–274. Dordrecht, Netherlands: Kluwer Acad.

Kuhl PK .(2000). A new view of language acquisition. *Proceedings of the National Academy of Sciences of the United States of America*, 97:11850-11857.

Kuhl PK, Miller JD. (1975). Speech perception by the chinchilla: voiced-voiceless distinction in alveolar plosive consonants. *Science*, 190:69–72.

Kuhl PK, Miller JD. (1978). Speech perception by the chinchilla: identification functions for synthetic VOT stimuli. *Journal of the Acoustical Society of America*, 63:905–917.

Liberman A, Mattingly IG. (1985). The motor theory of speech perception revised. *Cognition*, 21:1-36.

Liberman AM, Cooper FS, Shankweiler DP, Studdert-Kennedy M. (1967). Perception of the speech code. *Psychological Review*, 74:431-461.

Liberman AM, Whalen DH. (2000). On the relation of speech to language. *Trends in Cognitive Science*, 4:187–196.

Martin BA, Stapells DR. (2005). Effects of low-pass noise masking on auditory event-related potentials to speech. *Ear Hear*, 26:195-213.

Massaro DW. (1998). *Perceiving talking faces*. Cambridge, Massachusetts: MIT Press.

McEvoy L, Mäkelä JP, Hämäläinen M, Hari R. (1994). Effect of interaural time differences on middle-latency and late auditory evoked magnetic fields. *Hearing Research*, 78:249–257.

McGurk H, MacDonald J. (1976). Hearing lips and seeing voices. *Nature*, 264:746-748.

Meister IG, Wilson SM, Deblieck C, Wu AD, Iacoboni M. (2007). The essential role of premotor cortex in speech perception. *Current Biology*, 17(19):1692–1696.

Miettinen I, Tiitinen H, Alku P, May PJ. (2010). Sensitivity of the human auditory cortex to acoustic degradation of speech and non-speech sounds. BMC Neuroscience, 11:24.

Miller JD, Wier CC, Pastore RE, Kelly WJ, Dooling RJ. (1976). Discrimination and labeling of noise-buzz sequences with varying noise-lead times: an example of categorical perception. *Journal of the Acoustical Society of America*, 60:410–417.

Morita T, Nobuya F, Nagamine T, Hiraumi H, Naito Y, Shibasaki H, Ito J. (2006). Effects of continuous masking noise on tone-evoked magnetic fields in humans. *Brain Research*, 1087:151-158.

Mosher J, Leahy M. (1999). Source Localization Using Recursively Applied and Projected (RAP) MUSIC. *IEEE Transactions on signal processing*, 47:332-340.

Mosher JC, Lewis PS, Leahy RM. (1992). Multiple dipole modeling and localization from spatio-temporal MEG data. *IEEE Transactions on Biomedical Engineering*, 39:541-557.

Muller-Gass A, Marcoux A, Logan J, Campbell KB. (2001). The intensity of masking noise affects the mismatch negativity to speech sounds in human subjects. *Neuroscience Letters*, 299:197-200.

Murphy K, Corfield DR, Guz A, Fink GR, Wise RJ, Harrison J, Adams L. (1997). Cerebral areas associated with motor control of speech in humans. *Journal of Applied Physics*, 83:1438-1447.

Mäkelä JP, Hämäläinen M, Hari R, McEvoy L. (1994). Whole-head mapping of middle-latency auditory evoked magnetic fields. *Electroencephalography and Clinical Neurophysiology*; 92:414–421.

Narain C, Scott SK, Wise RJ, Rosen S, Leff A, Iversen SD, Matthews PM. (2003). Defining a left-lateralized response specific to intelligible speech using fMRI. *Cerebral Cortex*, 13:1362-1368.

Nishitani N, Hari R. (2000). Temporal dynamics of cortical representation for action. *Proceedings of the National Academy of Sciences of the United States of America*, 97:913-918.

Nishitani N, Hari R. (2002). Viewing lip forms: cortical dynamics. *Neuron*, 36:1211-1220.

Numminen J, Curio G. (1999). Differential effects of overt, covert and replayed speech on vowel-evoked responses of the human auditory cortex. *Neuroscience Letters*, 272(1):29-32.

Numminen J, Salmelin R, Hari R. (1999). Subject's own speech reduces reactivity of the human auditory cortex. *Neuroscience Letters*, 265:119-122.

Näätänen R, Picton T. (1987). The N1 wave of the human electric and magnetic response to sound: a review and an analysis of the component structure. *Psychophysiology*, 24(4):375-425.

Ojanen V, Pekkola J, Jääskeläinen IP, Möttönen R, Autti T, Jousmäki V, Sams M. (2005). *Common brain areas activated by hearing and seeing speech*. Laboratory of Computational Engineering Technical Report B49, ISBN 951-22-7679-8.

Ojemann G, Mateer C. (1979). Human language cortex: Localization of memory, syntax, and sequential motor-phoneme identification systems. *Science*, 205:1401-1403.

Okada Y. (1983). Neurogenesis of evoked magnetic fields. In: Williamson S, Romani G, Kaufman L, Modena L (editors). *Biomagnetism: an Interdisciplinary Approach*. pp. 399-408. New York and London: Plenum Press.

Okada YC, Wu J, Kyuhou S. (1997). Genesis of MEG signals in a mammalian CNS structure. *Electroencephalography and Clinical Neurophysiology*, 103:474-485.

Pang EW, Drake JM, Otsubo H, Martineau A, Strantzas S, Cheyne D, Gaetz W. (2008). Intraoperative confirmation of hand motor area identified preoperatively by magnetoencephalography. *Pediatric Neurosurgery*, 44(4):313-317.

Pekkola J, Ojanen V, Autti T, Jääskelainen IP, Möttönen R, Tarkiainen A, Sams M. (2005). Primary auditory cortex activation by visual speech: an fMRI study at 3 T. *Neuroreport*, 16(2):125-128.

Penfield W, Roberts L. (1959). Speech and Brain Mechanisms. Princeton University Press, Princeton, NJ.

Pickett E R, Kuniholm E, Protopapas A, Friedman J, Lieberman P. (1998). Selective speech motor, syntax and cognitive deficits associated with bilateral damage to the putamen and the head of the caudate nucleus: A case study. *Neuropsychologia*, 36:173-188.

Picton TW, Hillyard SA, Krausz HI, Galambos R. (1974). Human auditory evoked potentials. I. Evaluation of components. *Electroencephalography and Clinical Neurophysiology*, 36:179–190.

Pisoni DB. (1977). Identification and discrimination of the relative onset time of two component tones: implications for voicing perception in stops. *Journal of the Acoustical Society of America*, 61:1352-1361.

Pulvermüller F, Huss M, Kherif F, Moscoso del Prado Martin F, Hauk O, Shtyrov Y. (2006). Motor cortex maps articulatory features of speech sounds. *Proceedings of the National Academy of Sciences of the United States of America*, 103:7865-7870.

Rademacher J, Caviness VS, Jr., Steinmetz H, Galaburda AM. (1993). Topographical variation of the human primary cortices: implications for neuroimaging, brain mapping, and neurobiology. *Cerebral Cortex*, 3: 313-329.

Reed CM, Doherty MJ, Braida LD, Durlach NI. (1982). Analytic study of the Tadoma method: Further experiments with inexperienced observers. *Journal of Speech and Hearing Research*, 25:216-223.

Reisberg D, McLean J, Goldfield A. (1987). Easy to hear but hard to understand: A lipreading advantage with intact auditory stimuli. In: Dodd B, Campbell R (editors). *Hearing by Eye: The Psychology of Lip-reading*. pp. 97-113. London: Lawrence Erlbaum Associates.

Robinson SE, Vrba J. (1997). Functional neuroimaging by Synthetic Aperture Magnetometry (SAM). In: Yoshimoto T, Kotani M, Kuriki S, Karibe H, Nakasato B (editors). *Recent advances in biomagnetism*. pp. 302-305. Sendai: Tohoku University Press.

Romero L, Walsh V, Papagno C. (2006). The neural correlates of phonological short-term memory: A repetitive transcranial magnetic stimulation study. *Journal of Cognitive Neuroscience*, 18(7):1147–1155.

Rossing TD, Moore FR, Wheeler PA. (2002). *The Science of Sound*. San Francisco, London: Addison Wesley.

Roy AC, Craighero L, Fabbri-Destro M, Fadiga L. (2008). Phonological and lexical motor facilitation during speech listening: A transcranial magnetic stimulation study. Journal of Physiology, 102(1–3):101–105.

Salmelin R, Schnitzler A, Schmitz F, Freund H-J. (2000). Single word reading in developmental stutterers and fluent speakers. *Brain*, 123:1184-1202.

Sams M, Aulanko R, Hämäläinen M, Hari R, Lounasmaa OV, Lu ST, Simola J. (1991). Seeing speech: visual information from lip movements modifies activity in the human auditory cortex. *Neuroscience Letters*, 127:141–145

Sato M, Buccino G, Gentilucci M, Cattaneo L. (2010). On the tip of the tongue: modulation of the primary motor cortex during audiovisual speech perception. *Speech* Communication, 52(6): 533-554.

Sato M, Tremblay P, Gracco V. (2009). A mediating role of the premotor cortex in phoneme segmentation. *Brain and Language*, 111(1): 1-7.

Schafer EW, Amochaev A, Russell MJ. (1981). Knowledge of stimulus timing attenuates human evoked cortical potentials. *Electroencephalography and Clinical Neurophysiology*, 52(1):9-17.

Schafer EW, Marcus MM. (1973). Self-stimulation alters human sensory brain responses. *Science*. 181(95):175-177.

Scherg M, Vajsar J, Picton TW. (1989). A source analysis of the late human auditory evoked potentials. *Journal of Cognitive Neuroscience*, 1:336-355.

Schwartz J-L, Abry C, Boë L-J, Cathiard MA. (2002). Phonology in a theory of perception-for-action-control. In: Durand J, Lacks B (editors). *Phonology: From phonetics to cognition*. pp. 240-280. Oxford: Oxford University Press.

Schwartz J, Basirat A, Ménard L, Sato M. (in press). The Perception for Action Control Theory (PACT): a perceptuo-motor theory of speech perception. *Journal of Neurolinguistics*.

Schwartz JL, Boë LJ, Abry C. (2007). Linking the Dispersion-Focalization Theory (DFT) and the Maximum Utilization of the Available Distinctive Features (MUAF) principle in a Perception-for-Action-Control Theory (PACT). In: Solé MJ, Beddor P, Ohala M (editors), *Experimental Approaches to Phonology*. pp. 104-124. Oxford University Press.

Schwartz JL, Sato M, Fadiga L. (2008). The common language of speech perception and action: a neurocognitive perspective. *Revue Française de Linguistique Appliquée*, 13(2): 9-22.

Scott SK, Blank CC, Rosen S, Wise RJ. (2000). Identification of a pathway for intelligible speech in the left temporal lobe. *Brain*, 123:2400-2406.

Sekihara K, Nagarajan SS, Poeppel D, Marantz A, Miyashita Y. (2001). Reconstructing spatio-temporal activities of neural sources using an MEG vector beamformer technique. *IEEE Transactions on Biomedical Engineering*, 48:760-771.

Shtyrov Y, Kujala T, Ahveninen J, Tervaniemi M, Alku P, Ilmoniemi RJ, Näätänen R. (1998). Background acoustic noise and the hemispheric lateralization of speech processing in the human brain: magnetic mismatch negativity study. *Neuroscience Letters*, 2512:141-144.

Silveri MC, Di Betta AM, Filippini V, Leggio MG, Molinari M. (1998). Verbal short-term store-rehearsal system and the cerebellum: Evidence from a patient with a right cerebellar lesion. *Brain*, 121:2175-2177.

Skipper JI, Nusbaum HC, Small SL. (2005). Listening to talking faces: Motor cortical activation during speech perception. *NeuroImage*, 25(1):76-89.

Skipper JI, van Wassenhove V, Nusbaum HC, Small SL. (2007). Hearing Lips and Seeing Voices: How Cortical Areas Supporting Speech Production Mediate Audiovisual Speech Perception. *Cerebral Cortex*, 17:2387-2399.

Speech perception. (2010). In: Wikipedia, the free encyclopedia. Retrieved May 3, 2010, from http://en.wikipedia.org/wiki/Speech_perception

Spreng M. (1980). Influence of impulsive and fluctuating noise upon physiological excitations and short-time readaptation. *Scandinavian Audiology. Supplement*, 12:299-306.

Stam CJ, de Haan W, Daffertshofer A, Jones BF, Manshanden I, van Cappellen van Walsum AM, Montez T, Verbunt JP, de Munck JC, van Dijk BW, Berendse HW, Scheltens P. (2009). Graph theoretical analysis of magnetoencephalographic functional connectivity in Alzheimer's disease. *Brain*, 132(Pt1):213-224.

Stevens KN, Klatt DH. (1974). Role of formant transitions in the voiced-voiceless distinction for stops. *Journal of the Acoustical Society of America*, 55:653-659.

Stoffers D, Bosboom JL, Deijen JB, Wolters ECh, Stam CJ, Berendse HW. (2008). Increased cortico-cortical functional connectivity in early-stage Parkinson's disease: an MEG study. *Neuroimage*, 41(2):212-222.

Sumby WH, Pollack I. (1954). Visual contribution to speech intelligibility in noise. *Journal of the Acoustical Society of America*, 26:212-215.

Sundara M, Namasivayam AK, Chen R. (2001). Observation-execution matching system for speech: a magnetic stimulation study. *Neuroreport*, 12:1341-1344.

Tuomisto T, Hari R, Katila T, Poutanen T, Varpula T. (1983). Studies of auditory evoked magnetic and electric responses: Modality specificity and modelling. *Nuovo Cimento D*, 2:471-483.

Uutela K, Hämäläinen M, Somersalo E. (1999). Visualization of magnetoencephalographic data using minimum current estimates. *Neuroimage*, 10:173-180.

van Veen BD, Buckley K. (1988). Beamforming: A Versatile Approach to Spatial Filtering. *IEEE ASSP Magazine*, 5:4-24.

van Veen BD, van Drongelen W, Yuchtman M, Suzuki A. (1997). Localization of brain electrical activity via linearly constrained minimum variance spatial filtering. *IEEE Transactions on Biomedical Engineering*, 44:867-880.

von Helmholtz Hv. (1853). Über einige Gesetze der Vertheilung elektrischer Ströme in körperlichen Leitern, mit Anwendung auf die thierisch-elektrischen Versuche. *Annual Review of Physical Chemistry*, 89:211-233, 353-377.

Vouloumanos A, Kiehl KA, Werker JF, Liddle PF. (2001). Detection of sounds in the auditory stream: event-related fMRI evidence for differential activation to speech and nonspeech. *Journal of Cognitive Neuroscience*, 13:994-1005.

Watkins K, Paus T. (2004). Modulation of motor excitability during speech perception: the role of Broca's area. *Journal of Cognitive Neuroscience*, 16:978-987.

Watkins KE, Strafella AP, Paus T. (2003). Seeing and hearing speech excites the motor system involved in speech production. *Neuropsychologia*, 41:989-994.

Wernicke C. (1874). Der Aphasische Symptomencomplex [The aphasia symptom complex]. Cohn and Weigert, Breslau.

Whiting KA, Martin BA, Stapells DR. (1998). The effect of broadband noise masking on cortical event-related potentials to speech sounds /ba/ and /da/. *Ear Hear*, 19:218-231.

Wise RJS, Greene J, Buchel C, Scott SK. (1999). Brain regions involved in articulation. *Lancet*, 353:1057-1061.

Williams RW, Herrup K. (1988). The control of neuron number. *Annual Review of Neuroscience*, 11:423-453.

Wilson SM, Iacoboni M. (2006). Neural responses to non-native phonemes varying in producibility: Evidence for the sensorimotor nature of speech perception. *NeuroImage*, 33(1):316-325.

Wilson SM, Saygin AP, Sereno MI, Iacoboni M. (2004). Listening to speech activates motor areas involved in speech production. *Nature Neuroscience*, 7:701-702.

Yoshiura T, Ueno S, Iramina K, Masuda K, Kiyuna T. (1995). Measurement and analysis of middle latency auditory evoked magnetic fields. In: Baumgartner C, Deecke L, Stroink G, Williamson SJ (Editors), *Biomagnetism: Fundamental Research and Clinical Applications*. pp. 205-209. Amsterdam: Elsevier Science, IOS Press.

Zekveld AA, Heslenfeld DJ, Festen JM, Schoonhoven R. (2006). Top–down and bottom–up processes in speech comprehension. *Neuroimage*, 32:1826–1836.

Ziegler W, Kilian B, Deger K. (1997). The role of the left mesial frontal cortex in fluent speech: Evidence from a case of left supplementary motor area hemorrhage. *Neuropsychologia*, 35:1197-1208.

Zimmerman JE, Silver AH. (1966). Macroscopic quantum interference effects through superconducting point contacts. *Physical Review*, 141:367-375.

Zouridakis G, Simos PG, Papanicolaou AC. (1998). Multiple bilaterally asymmetric cortical sources account for the auditory N1m component. *Brain topography*, 10(3):183-189.

Appendix A

Presentation scripts used in the study

A.1 Example script for the preliminary discrimination tests (noises1.sce)

```
scenario = "Noises1";
no_logfile = true;
active_buttons = 3;
button_codes = 1, 2, 0;       /* Codes 1 and 2 are for
                              responses for /pa/ and /ta/,
                              respectively.
                              Code 0 is for 'Enter', which
                              starts the scenario. */
target_button_codes = 11, 12, 0;        /* Different codes for
                                        correct responses */
response_matching = simple_matching;
response_logging = log_active; # No entries in the logfile
before the fixation cross turns red
default_background_color = 255, 255, 255;
$black = "0, 0, 0";
$red = "255, 0, 0";

begin;

picture {} default;

picture { # black fixation cross
      text {
                    caption = "+";
                    font_color = $black;
          font_size = 200;
      };
      x = 0; y = 0;
   } black_fixation_cross;

picture { # red fixation cross
          text {
                    caption = "+";
                    font_color = $red;
          font_size = 200;
      };
      x = 0; y = 0;
   } red_fixation_cross;

trial {
```

```
   trial_type = first_response;
   trial_duration = forever;

   picture {
      text {
          caption = "Kohinatavujen tunnistaminen 1";
          font_color = $black;
          font_size = 48;
      };
      x = 0; y = 120;
      text {
         caption =  "Näytön keskellä olevan ristin ollessa
                    musta kuulet joko /pa/- tai /ta/-tavun
                    kohinaan upotettuna. Ristin muututtua
                    punaiseksi paina painiketta 'P', jos
                    kuulit /pa/ tai painiketta 'T', jos kuulit
                    /ta/. Annettuasi vastauksen risti muuttuu
                    jälleen mustaksi ja kuulet seuraavan
                    tavun.
                    \nPaina 'Enter' aloittaaksesi.";
          font_color = $black;
          font_size = 24;
      };
      x = 0; y = -120;
   };
   time = 0;
} ;

TEMPLATE "noises.tem" randomize { # noise stimuli
          file                event_code          target_code;
          "1_pa_noise1.wav"   "1_pa_noise1"       1          ;
          "1_ta_noise1.wav"   "1_ta_noise1"       2          ;
          "2_pa_noise1.wav"   "2_pa_noise1"       1          ;
          "2_ta_noise1.wav"   "2_ta_noise1"       2          ;
          "3_pa_noise1.wav"   "3_pa_noise1"       1          ;
          "3_ta3_noise1.wav"  "3_ta_noise1"       2          ;
          "4_pa_noise1.wav"   "4_pa_noise1"       1          ;
          "4_ta_noise1.wav"   "4_ta_noise1"       2          ;
          "5_pa_noise1.wav"   "5_pa_noise1"       1          ;
          "5_ta_noise1.wav"   "5_ta_noise1"       2          ;
          "1_pa_noise1.wav"   "1_pa_noise1"       1          ;
          "1_ta_noise1.wav"   "1_ta_noise1"       2          ;
          "2_pa_noise1.wav"   "2_pa_noise1"       1          ;
          "2_ta_noise1.wav"   "2_ta_noise1"       2          ;
          "3_pa_noise1.wav"   "3_pa_noise1"       1          ;
          "3_ta_noise1.wav"   "3_ta_noise1"       2          ;
          "4_pa_noise1.wav"   "4_pa_noise1"       1          ;
          "4_ta_noise1.wav"   "4_ta_noise1"       2          ;
          "5_pa_noise1.wav"   "5_pa_noise1"       1          ;
          "5_ta_noise1.wav"   "5_ta_noise1"       2          ;
          "1_pa_noise1.wav"   "1_pa_noise1"       1          ;
          "1_ta_noise1.wav"   "1_ta_noise1"       2          ;
          "2_pa_noise1.wav"   "2_pa_noise1"       1          ;
```

```
            "2_ta_noise1.wav"   "2_ta_noise1"       2          ;
            "3_pa_noise1.wav"   "3_pa_noise1"       1          ;
            "3_ta_noise1.wav"   "3_ta_noise1"       2          ;
            "4_pa_noise1.wav"   "4_pa_noise1"       1          ;
            "4_ta_noise1.wav"   "4_ta_noise1"       2          ;
            "5_pa_noise1.wav"   "5_pa_noise1"       1          ;
            "5_ta_noise1.wav"   "5_ta_noise1"       2          ;
            "1_pa_noise1.wav"   "1_pa_noise1"       1          ;
            "1_ta_noise1.wav"   "1_ta_noise1"       2          ;
            "2_pa_noise1.wav"   "2_pa_noise1"       1          ;
            "2_ta_noise1.wav"   "2_ta_noise1"       2          ;
            "3_pa_noise1.wav"   "3_pa_noise1"       1          ;
            "3_ta_noise1.wav"   "3_ta_noise1"       2          ;
            "4_pa_noise1.wav"   "4_pa_noise1"       1          ;
            "4_ta_noise1.wav"   "4_ta_noise1"       2          ;
            "5_pa_noise1.wav"   "5_pa_noise1"       1          ;
            "5_ta_noise1.wav"   "5_ta_noise1"       2          ;
            "1_pa_noise1.wav"   "1_pa_noise1"       1          ;
            "1_ta_noise1.wav"   "1_ta_noise1"       2          ;
            "2_pa_noise1.wav"   "2_pa_noise1"       1          ;
            "2_ta_noise1.wav"   "2_ta_noise1"       2          ;
            "3_pa_noise1.wav"   "3_pa_noise1"       1          ;
            "3_ta_noise1.wav"   "3_ta_noise1"       2          ;
            "4_pa_noise1.wav"   "4_pa_noise1"       1          ;
            "4_ta_noise1.wav"   "4_ta_noise1"       2          ;
            "5_pa_noise1.wav"   "5_pa_noise1"       1          ;
            "5_ta_noise1.wav"   "5_ta_noise1"       2          ;

};
```

A.2 Supplementary template for script A.1 (noises.tem)

```
trial {
   all_responses = false; # Ignore responses that occur before
the fixation cross turns red
   trial_type = first_response;
   trial_duration = forever;

   picture black_fixation_cross;
      time = 0;
      duration = 1600;

   sound {
      wavefile { filename = $file; };
      };
      time = 1200;
      code = $event_code;
      port_code = $output_code;

picture red_fixation_cross;
```

```
   time = 1600;
   target_button = $target_code;        /* Correct response
                                        gives a different code
                                        */
};
```

A.3 Example script for the main experiment (passive.sce)

```
scenario = "Passive perception";
active_buttons = 1;
button_codes = 1;
write_codes = true; # Codes are written to the output port
when an event occurs
pulse_width = 20; # if port is parallel
default_background_color = 255, 255, 255;
$black = "0, 0, 0";
$red = "255, 0, 0";
$cross_size = 300;

begin;

picture { # This is for eliminating the vertical blanking
period between trials
          text {
            caption = "+";
            font_color = $black;
            font_size = $cross_size;
      };
      x = 0; y = 0;
          } default;


# Black fixation cross
picture {
      text {
            caption = "+";
            font_color = $black;
            font_size = $cross_size;
      };
      x = 0; y = 0;
   } black_fixation_cross;


# Red fixation cross
picture {
                    text {
            caption = "+";
            font_color = $red;
            font_size = $cross_size;
```

```
      };
      x = 0; y = 0;
   } red_fixation_cross;


# Stimuli
array {
         TEMPLATE "sound.tem" {
                  file                                    ;
/* 1 */           "1_pa300ms_68.wav"                      ;
/* 2 */           "1_ta300ms_68.wav"                      ;
/* 3 */           "2_pa300ms_68.wav"                      ;
/* 4 */           "2_ta300ms_68.wav"                      ;
/* 5 */           "3_pa300ms_68.wav"                      ;
/* 6 */           "3_ta300ms_68.wav"                      ;
/* 7 */           "4_pa300ms_68.wav"                      ;
/* 8 */           "4_ta300ms_68.wav"                      ;
/* 9 */           "5_pa300ms_68.wav"                      ;
/* 10 */          "5_ta300ms_68.wav"                      ;
/* 11 */          "1_pa300ms_noise1.wav"                  ;
/* 12 */          "1_ta300ms_noise1.wav"                  ;
/* 13 */          "2_pa300ms_noise1.wav"                  ;
/* 14 */          "2_ta300ms_noise1.wav"                  ;
/* 15 */          "3_pa300ms_noise1.wav"                  ;
/* 16 */          "3_ta300ms_noise1.wav"                  ;
/* 17 */          "4_pa300ms_noise1.wav"                  ;
/* 18 */          "4_ta300ms_noise1.wav"                  ;
/* 19 */          "5_pa300ms_noise1.wav"                  ;
/* 20 */          "5_ta300ms_noise1.wav"                  ;
                  };
} stimuli;



# Instruction trial; in Finnish
trial {
   trial_type = first_response;
   trial_duration = forever;

   picture {
      text {
         caption = "Passiivinen havaitseminen";
         font_color = $black;
                              font_size = 48;
      };
                  x = 0; y = 100;
      text {
         caption =
"Seuraavassa kokeessa näet näytön keskellä mustan ristin ja
kuulet puhuttuja 'pa'- ja 'ta'-tavuja. Pidä katse lukittuna
ristin keskiosaan. Jokaisen tavun jälkeen ohjelma antaa
sinulle merkin vilauttamalla ristiä punaisena.
Koska kyseessä on passiivinen havaitseminen,
```

```
sinun ei tarvitse tehdä mitään tämän merkin nähtyäsi.";

         font_color = $black;
         font_size = 24;
      };
      x = 0; y = -100;
   };
   time = 0;
} instruction_trial;


/*
# Instruction trial; in English
trial {
   trial_type = first_response;
   trial_duration = forever;

   picture {
      text {
         caption = "Passive perception";
         font_color = $black;
                              font_size = 48;
      };
                  x = 0; y = 120;
      text {
         caption =
"In the next scenario, you will see
a black cross at the center of the screen
and hear spoken 'pa' and 'ta' syllables.
After each syllable, the program gives you
a sign by turning the color of the cross briefly to red.
However, as this is passive perception, you don't need
to do anything as you see this sign.
\nPress 'Enter' to start.";

         font_color = $black;
         font_size = 24;
      };
      x = 0; y = -120;
   };
   time = 0;
} instruction_trial;
*/


# Main trial
trial {
   picture black_fixation_cross;
   time = 0;
   duration = 2000;

           stimulus_event {    /* This represents the auditory
                               stimulus. It is accessed in the
```

```
                                PCL section, where also the
                                actual sound stimulus and all
                                the parameters, such as the
                                stimulus onset time, are
                                assigned. */
            nothing {};
    } event;

            picture red_fixation_cross;
            time = 2000;
            duration = 200;
            port_code = 21;

            picture black_fixation_cross;
            time = 2200;
            duration = 3800;
} main_trial;


begin_pcl;


# Event codes
array <string> event_codes[stimuli.count()] = {
/* 1 */             "1_pa300ms_68.wav",
/* 2 */             "1_ta300ms_68.wav",
/* 3 */             "2_pa300ms_68.wav",
/* 4 */             "2_ta300ms_68.wav",
/* 5 */             "3_pa300ms_68.wav",
/* 6 */             "3_ta300ms_68.wav",
/* 7 */             "4_pa300ms_68.wav",
/* 8 */             "4_ta300ms_68.wav",
/* 9 */             "5_pa300ms_68.wav",
/* 10 */            "5_ta300ms_68.wav",
/* 11 */            "1_pa300ms_noise1.wav",
/* 12 */            "1_ta300ms_noise1.wav",
/* 13 */            "2_pa300ms_noise1.wav",
/* 14 */            "2_ta300ms_noise1.wav",
/* 15 */            "3_pa300ms_noise1.wav",
/* 16 */            "3_ta300ms_noise1.wav",
/* 17 */            "4_pa300ms_noise1.wav",
/* 18 */            "4_ta300ms_noise1.wav",
/* 19 */            "5_pa300ms_noise1.wav",
/* 20 */            "5_ta300ms_noise1.wav"
};


# Port codes
array <int> port_codes[stimuli.count()] = {
/* 1 */   1,
/* 2 */   2,
/* 3 */   3,
/* 4 */   4,
```

```
/* 5 */   5,
/* 6 */   6,
/* 7 */   7,
/* 8 */   8,
/* 9 */   9,
/* 10 */  10,
/* 11 */  11,
/* 12 */  12,
/* 13 */  13,
/* 14 */  14,
/* 15 */  15,
/* 16 */  16,
/* 17 */  17,
/* 18 */  18,
/* 19 */  19,
/* 20 */  20
};

array <int> tracker[stimuli.count()];   /* Used to keep track
                                           on the number each
                                           stimulus has been
                                           presented */
int number_of_trials = 150;  # Number of trials


instruction_trial.present();


loop
   int i = 1
until
   i > number_of_trials
begin

          int index = random( 1, stimuli.count() );
          /* A number between 1 and the number of stimuli
          used is randomly generated */

/* The if-then construct below checks whether the randomly
picked stimulus has already occurred the number of times it is
allowed to for even distribution. If it has, then the
execution of the loop will start from the beginning, i.e.
'index' is assigned a new randomly picked value. */
if (tracker[index] > ( number_of_trials / stimuli.count() - 1
)) then
continue;           /* The continue statement causes the
                    remainder of the loop body to be skipped;
                    however, it does not exit the loop. */
end;

event.set_stimulus( stimuli[index] );   /* Sound stimulus is
                                           set for the empty
                                           stimulus_event */
```

```
event.set_time( random( 1000, 1500 ) ); /* The onset time for
                                        the sound stimulus is
                                        randomly generated
                                        between 1000 and 1500
                                        ms. */
# Event code isgiven to the stimulus.
event.set_event_code( event_codes[index] );

# Port code is given to the stimulus.
event.set_port_code( port_codes[index] );

main_trial.present();

tracker[index] = tracker[index] + 1;
# Tracker array update
i = i + 1

end;
```

A.4 Supplementary template for script A.3 (sound.tem)

```
sound {
          wavefile { filename = $file; };
};
```