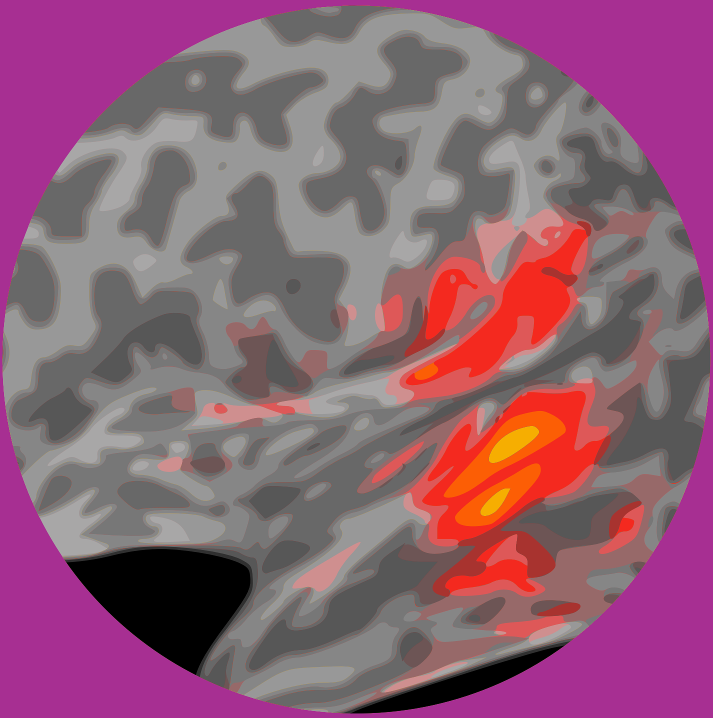
Department of Biomedical Engineering and Computational Science

Measures and models of top-down influences in the human auditory cortex

Jaakko Kauramäki



Aalto University

DOCTORAL DISSERTATIONS

Aalto University publication series
DOCTORAL DISSERTATIONS 58/2012

Measures and models of top-down influences in the human auditory cortex

Jaakko Kauramäki

Doctoral dissertation for the degree of Doctor of Science in Technology to be presented with due permission of the School of Science for public examination and debate in Auditorium of Department of Biomedical Engineering and Computational Science at the Aalto University School of Science (Espoo, Finland) on the 16th of May 2012 at 12 noon.

Aalto University
School of Science
Department of Biomedical Engineering and Computational Science
Brain and Mind Laboratory

Supervisor
Professor Mikko Sams

Instructor
Docent Iiro Jääskeläinen

Preliminary examiners
Professor Howard Nusbaum, University of Chicago, USA
Professor Kimmo Alho, University of Helsinki, Finland

Opponent
Professor Jyrki Ahveninen, Harvard Medical School, USA

Aalto University publication series
DOCTORAL DISSERTATIONS 58/2012

ISBN 978-952-60-4619-8 (printed)
ISBN 978-952-60-4620-4 (pdf)
ISSN-L 1799-4934
ISSN 1799-4934 (printed)
ISSN 1799-4942 (pdf)

Unigrafia Oy
Helsinki 2012

Finland

441 697
Printed matter

Aalto University

Aalto University, P.O. Box 11000, FI-00076 Aalto www.aalto.fi

Abstract

Author
Jaakko Kauramäki

Name of the doctoral dissertation
Measures and models of top-down influences in the human auditory cortex

Publisher School of Science

Unit Department of Biomedical Engineering and Computational Science

Series Aalto University publication series DOCTORAL DISSERTATIONS 58/2012

Field of research Cognitive neuroscience

Manuscript submitted 24 January 2012 **Manuscript revised** 5 April 2012

Date of the defence 16 May 2012 **Language** English

☐ **Monograph** ☒ **Article dissertation (summary + original articles)**

Abstract

Our senses rapidly adapt to the sensory information and task requirements. These short-term, plastic changes especially in the central nervous system are necessary for goal-directed and active behavior. Nevertheless, partly due to methodological limitations, the neural mechanisms underlying improved perception during selective attention are not well understood. For instance, human auditory cortex activity is enhanced while engaging in an auditory task, but whether such enhancement involves a larger-extend or a more specific response from a sharply tuned neuronal population remains vague. In addition to unimodal attention effects, our senses can influence each others' processing. Seeing a person articulate can both enhance and modify the perception of acoustic speech and even induce activation in the auditory areas. The goal of the Thesis was to study the neural basis of these auditory cortex modulations both due to selective attention and cross-modal interactions while lipreading.

This Thesis consists of a series of four studies where the brain activity was recorded with electroencephalography (EEG) and its magnetic counterpart, magnetoencephalography (MEG). The results show that 1) evoked responses mainly from the secondary auditory areas were rapidly enhanced by a feature-specific manner during a discrimination task, where bottom-up and top-down processing were modified by stimulus and task design, 2) lipreading task suppressed auditory-cortex responses both generally and specifically in the case when sound was related to the articulatory gestures, and 3) lipreading and covert speech production tasks dynamically suppressed auditory cortex reactivity, possibly via an efference copy signal from the speech-production system.

Together, these studies expand the knowledge on how the human auditory cortex function is modulated under top-down influence. The modulations due to selective attention and cross-modal interactions while lipreading are opposite at the level of evoked responses: selective attention enhances and lipreading suppresses the responses, but both effects can be feature-selective. The results show that the auditory system is highly dependent on the current task requirements, showing adaptive and goal-dependent functioning.

Keywords electroencephalography, magnetoencephalography, top-down, attention, lipreading, human brain, auditory system, perception

ISBN (printed) 978-952-60-4619-8 **ISBN (pdf)** 978-952-60-4620-4

ISSN-L 1799-4934 **ISSN (printed)** 1799-4934 **ISSN (pdf)** 1799-4942

Location of publisher Espoo **Location of printing** Helsinki **Year** 2012

Pages 130 **The dissertation can be read at** http://lib.tkk.fi/Diss/

Aalto-yliopisto

Tiivistelmä

Aalto-yliopisto, PL 11000, 00076 Aalto www.aalto.fi

Tekijä
Jaakko Kauramäki

Väitöskirjan nimi
Mittauksia ja malleja ihmisen kuuloaivokuoren top-down -vaikutuksista

Julkaisija Perustieteiden korkeakoulu

Yksikkö Lääketieteellisen tekniikan ja laskennallisen tieteen laitos

Sarja Aalto University publication series DOCTORAL DISSERTATIONS 58/2012

Tutkimusala Kognitiivinen neurotiede

Käsikirjoituksen pvm 24.01.2012 **Korjatun käsikirjoituksen pvm** 05.04.2012

Väitöspäivä 16.05.2012 **Kieli** Englanti

☐ **Monografia** ☒ **Yhdistelmäväitöskirja (yhteenveto-osa + erillisartikkelit)**

Tiivistelmä

Ihmisen aistit mukautuvat nopeasti havaintoinformaation ja tilanteen vaatimusten mukaan. Nämä lyhytaikaiset, plastiset muutokset erityisesti keskushermoston tasolla ovat välttämättömiä tehokkaalle toiminnalle. Tarkkaavaisuustilanteissa todetun parantuneen havaintokyvyn hermostollista perustaa ei kuitenkaan täysin tunneta, osin rajoitettujen tutkimusmenetelmien takia. Esimerkiksi ihmisen kuuloaivokuoren toiminta tehostuu ääneen liittyvää tehtävää tehdessä, mutta on epäselvää liittyykö tähän tehostumiseen hermosolujen laajempialaiset vai tarkemmin virittyneet aktivaatiot. Yhteen aistiin liittyvien tehostumismekanismien lisäksi aistit vaikuttavat toinen toistensa toimintaan. Toisen puhujan huulten liikkeiden näkeminen aktivoi kuuloaivokuorta, sen lisäksi huuliltaluku sekä helpottaa puheen ymmärtämistä että tietyissä tilanteissa muuttaa havaintoa. Tämän väitöskirjatyön tavoitteena oli tutkia valikoivan tarkkaavaisuuden ja huuliltaluvun aikaisen aistienvälisen vuorovaikutuksen hermostollista perustaa terveiden ihmisten kuuloaivokuorella.

Väitöskirjatyö koostuu neljästä osatutkimuksesta, joissa aivojen aktivaatioiden aiheuttamia sähkömagneettisia signaaleja tutkittiin käyttämällä tutkimusmenetelminä sekä elektroenkefalografiaa (EEG) että magnetoenkefalografiaa (MEG). Tulokset osoittivat, että 1) pääosin sekundaarisella kuuloaivokuorella syntyvät herätevasteet muovautuivat piirreriippuvaisesti sellaisen äänien erottelutehtävän aikana, jossa bottom-up ja top-down -prosessointia muokattiin monimutkaisilla ääniärsykkeillä ja ohjeistuksella, 2) huuliltaluku vaimensi kuuloaivokuorella syntyneitä vasteita sekä yleisesti että piirreriippuvaisesti mikäli ääni liittyi nähtyyn artikulaatioon, 3) huuliltaluku ja hiljainen puheentuotto vaimensivat dynaamisesti kuuloaivokuoren vasteita, mahdollisesti puheentuottojärjestelmästä lähteneen efferentin hermovasteen takia.

Tämän väitöskirjan tulokset laajentavat tietämystä siitä, miten top-down -vaikutukset näkyvät ihmisen kuuloaivokuorella. Tulosten perusteella voidaan sanoa, että kuuloaivokuoren toiminnassa näkyy valikoivan tarkkaavaisuuden ja huuliltalukutehtävän aikana suunnaltaan päinvastaisia muutoksia verrokkitilanteeseen nähden: tarkkaavaisuus suurentaa ja huuliltaluku vaimentaa vasteita siten, että molemmat vaikutukset ovat riippuvaisia käytetyistä ärsykkeistä. Tämä osoittaa, että ihmisen kuulojärjestelmän toiminta muovautuu nopeasti tehtävän vaatimusten mukaan.

Avainsanat elektroenkefalografia, magnetoenkefalografia, top-down, tarkkaavaisuus, huuliltaluku, ihmisaivot, kuulojärjestelmä, havaitseminen

ISBN (painettu) 978-952-60-4619-8 **ISBN (pdf)** 978-952-60-4620-4

ISSN-L 1799-4934 **ISSN (painettu)** 1799-4934 **ISSN (pdf)** 1799-4942

Julkaisupaikka Espoo **Painopaikka** Helsinki **Vuosi** 2012

Sivumäärä 130 **Luettavissa verkossa osoitteessa** http://lib.tkk.fi/Diss/

Preface

This work was carried out at the Department of Biomedical Engineering and Computational Science (BECS), Aalto University School of Science. I wish to express my gratitude to my supervisor, Professor Mikko Sams, and instructor, Docent Iiro Jääskeläinen, for the valuable and educating opportunity to work and finish my doctoral thesis under your guidance. Your scientific knowledge is vast and the standards you set exemplary. This period has been a great learning experience for me. I did not know much about auditory neuroscience back then when I started, years ago. I still feel that I do not know that much, but at least I now understand my limitations and aim to expand my knowledge, thanks to you.

This work was made possible by financial support from various sources. I am grateful to the Finnish Graduate School of Neuroscience (FGSN), Emil Aaltonen Foundation, Ella and Georg Ehrnrooth Foundation, and TKK Research Foundation, for granting me monetary support to enable full-time research for several years.

I would like to thank my preliminary examiners, Professors Howard C. Nusbaum and Kimmo Alho for the feedback on the summary of the Thesis that was both inspirating and constructive and helped to finalize the work.

The MEG data was collected at the O.V. Lounasmaa Laboratory of Aalto University, and MRI images used in the MEG studies were from Advanced Magnetic Imaging (AMI) Centre, Aalto University. I am grateful to the head of Brain Research Unit (BRU) of the O.V. Lounasmaa Laboratory, Academician Riitta Hari, for both letting me conduct research there and for spreading the enthusiastic attitude towards science, and personnel of BRU for the help in setting up the MEG measurements and analyses.

The publications included in this thesis were all the end result of a group effort. I would like to thank my co-authors in these four publications for making the results come to life. In addition, I would like to express my gratitude to my coworkers at BECS and its predecessor, Laboratory of Cognitive Engineering (LCE), for the inspirating company, nice overall atmosphere and good parties. Special warm hugs to people and hang-arounds of the Magnet House. You started as colleagues, but became dear friends even after leaving the academy or parting ways for other reasons. Something to do with the strong magnetic field aligning people along its field lines, I guess. Also, I would like to thank all my test subjects for their participation and patience. Volunteering to listen to filtered white noise for over an hour in a small chamber really shows some dedication to science. A warm thank you is in place also to the secretaries for handling the not-always-so-clever but mandatory bureacratic issues, and the IT support for fast help when needed.

I would like to thank my parents and siblings for the love, support and companionship during my younger years, allowing me to make my own choices in life. Special acknowledgment to my father and mother for the practical help during the recent years with our small children. Without that, finishing this project would probably have lasted even longer.

Finally, to my current family. I would like to thank my wife Jaana for all the love and understanding during the last decade and the decades to come. Finding a proper balance between the academic life and work, hobbies, and spending time with the family has not been easy, but you have rightfully, yet lovingly, reminded me on what is important in the long run. And, last but not least, I would like to thank my dearest three little princesses: Maj, Tiu and Uni, for spending countless enjoyable hours playing with yours truly and letting the father-daughter time enlighten my mood after the long work days. During the final steps of this Thesis process, you have learned to instantly recognize a schematic picture of a human brain from the printouts I have brought home, but most importantly you have learned to ask so hard questions on brain function that I cannot answer without checking some references. And, that is what we should all aim to in science, asking the right questions.

Espoo, April 23, 2012,

Jaakko Kauramäki

Contents

List of Publications

This thesis consists of an overview and of the following publications which are referred to in the text by their Roman numerals.

I **Kauramäki, J.**, Jääskeläinen, I. P., and Sams, M. Selective attention increases both gain and feature selectivity of the human auditory cortex. *PLoS ONE*, 2(9): e909, September 2007.

II **Kauramäki, J.**, Jääskeläinen I. P., Hänninen, J. L., Auranen, T., Nummenmaa, A., Lampinen, J., Sams, M. Dual-stage adaptive filtering of sounds during selective attention. *PLoS ONE*, Submitted, December 2011.

III Jääskeläinen, I. P., **Kauramäki, J.**, Tujunen, J., and Sams, M. Formant transition specific adaptation by lipreading of left auditory cortex N1m. *NeuroReport*, 19(1): 93–97, January 2008.

IV **Kauramäki, J.**, Jääskeläinen, I. P., Hari, R., Möttönen, R., Rauschecker, J. P., and Sams, M. Lipreading and covert speech production similarly modulate human auditory-cortex responses to pure tones. *Journal of Neuroscience*, 30(4): 1314–1321, January 2010.

Author’s Contribution

Publication I: “Selective attention increases both gain and feature selectivity of the human auditory cortex”

The author designed and prepared the experimental setup with other co-authors, collected the experimental data, analyzed the data and wrote the manuscript together with the co-authors.

Publication II: “Dual-stage adaptive filtering of sounds during selective attention”

The author designed and prepared the experimental setup with other co-authors, collected 30% of the experimental MEG data, analyzed the data and wrote the manuscript together with the co-authors.

Publication III: “Formant transition specific adaptation by lipreading of left auditory cortex N1m”

The author had a crucial role in the experimental setup design and programming, assisted in the MEG data collection, and had an active role in the data analysis and manuscript writing.

Publication IV: “Lipreading and covert speech production similarly modulate human auditory-cortex responses to pure tones”

As in Publication I.

List of Abbreviations

AC	Auditory cortex
ACh	Acetylcholine
BM	Basilar membrane
ECD	Equivalent current dipole
EEG	Electroencephalography
ERF / ERP	Event-related field / potential
FM	Frequency-modulated
fMRI	Functional magnetic resonance imaging
IC	Inferior colliculus
IHC	Inner hair cell
MEG	Magnetoencephalography
MGN	Medial geniculate nucleus
MNE	Minimum norm estimate
MRI	Magnetic resonance imaging
Nd	Negative difference
Ndm	Negative difference (for magnetic fields)
OHC	Outer hair cell
PF / PN	Processing field / negativity
RF	Radio-frequency or receptive field
SC	Superior colliculus
SQUID	Superconducting quantum interference device
STS	Superior temporal sulcus

1. Introduction

The human brain is responsible for deciphering the world around us, and to allow us to interact with the environment appropriately. The auditory system analyzes all the sounds, all the fine details included in them, the location where the sounds originate from, in an instant. It is responsible for the first steps in decoding the sound to a meaning, but unlike a mechanical machine, its functioning is not fixed. It is not a "black box" of which output is always determined by the input. The current state of the brain, how vigilant the person is, the objectives and mood of the person hearing a sound or listening to it, all affect how the sound is processed and what is the outcome. One example of this is that we are able to hear the nuances of a musical piece, the deliberate introduction of rich dynamics by the composer or the instrument player, by carefully listening to the musical piece. Or, when we are alone in a dark forest, we can hear the cracks and wind humming very vividly, attaching meaning even to the faintest of sounds, and a sudden unexpected sound can capture our attention in milliseconds. This is an example of both bottom-up and top-down attentional processes. A salient sound "pops out" and works as the bottom-up driver, capturing our attention and changing our behavior instantly. The amplification of faint sounds is an example of top-down influence, where a signal coming from higher-order areas of our brain, such as the prefrontal cortex, connects to the sensory system, and modulates the typical functioning of the system decoding the input from the outside world.

We can focus our attention to the incoming sound. We can choose who to listen to in a "cocktail party" (Cherry, 1953) setting where we are surrounded by a number of speakers in different locations, and ignore the others. There, we are able to voluntarily select another speaker to attend to, or rapidly switch the speaker grabbing our attention by mentioning our name, proving that the suppressed background noise is still processed at some level. Further, we can focus on the small details or features of the sound. We are capable of finding very small irregularities or oddities in the sound, for instance when finding for the cause of the malfunctioning car by listening the engine revving. Practise and previous exposure to similar sounds makes us even better, and an experienced car mechanic might be able to say right away what is the cause of the altered motor sound.

Long-term changes in the brain structures and auditory system functioning have been shown in musicians compared to nonmusicians (Münte et al., 2002). The short-term enhancements during selective attention, occuring in seconds or minutes after the task begins, are reflected in the brain activity as well. In a laboratory setting, it has been shown that choosing which ear to listen to, not letting it interfere with a competing message coming to the other ear, is reflected in the brain activity (Hillyard et al., 1973), even though we cannot voluntarily turn off the input to either ear. Further evidence of strong effects of actively listening are shown in the first two studies of this Thesis.

The senses can further interact with each other, a phenomena which is exploited in the two last studies of this Thesis. We can understand speech better if we see the speaker (Sumby and Pollack, 1954). The noisier the cocktail party, the more we rely on the visual lip movements when trying to follow the speaker. Seeing speech helps us to create internal predictions on what is the meaning of the word by reducing the ambivalence of speech sound. Speech sounds are not fixed, and the sound pattern of speech sound constituents, phonemes, vary depending on where the phoneme is in a word. This is called coarticulation (Fowler and Saltzman, 1993), and it further exemplifies how a straightforward sound signal analysis cannot explain all the functions the human auditory system is responsible of.

A lot of the sound processing is done in the auditory cortex, located deep in the folds of temporal cortex. This Thesis focuses on the cortical activity due to the measurement methods that were used. Changes in the lower regions are possible as well, and even the functioning of the inner ear, cochlea, is rapidly modified during various task conditions (Giard et al., 1994).

This Thesis presents four studies where the top-down influences introduced by task design is reflected in the brain activity. The aim of the first two studies was to investigate whether a “gain” model (Hillyard et al., 1998) explains the auditory selective attention, increasing the neural activity to all sounds, or does increase in feature selectivity additionally underlie the enhanced perception. The two last studies concentrate on how a lipreading task changes the auditory cortex function. Previous studies have shown that the auditory cortical responses are suppressed (Numminen and Curio, 1999; Curio et al., 2000). Publication III asks whether the suppression is general or specific to the lipread sound, and Publication IV asks how the suppression during lipreading task is related to the speech production system.

2. Background

2.1 Neurophysiology

2.1.1 Neural signaling

Neuronal communication takes place through action potentials traveling in the axons. Axons typically connect to dendrites of other neurons, but can also connect directly to soma (cell body) or even other axons. The interesting part, however, happens at the synapses between the axons and dendrites, where the signals are modified (for a review, Burrone and Murthy, 2003; Abbott and Regehr, 2004). Whether input signal passes on is based on the current state of the synapse, for instance, whether there has been recent synaptic input from other neurons. The synapse weights are modified constantly at different time scales by experience and learning. In simplest means this learning is Hebbian: the more often the neurons are active together, the stronger the connections will become.

In synapses, the communication in the brain is typically chemical, via the release of neurotransmitters. This release, in turn, strongly affects the ion concentration between the pre- and postsynaptic terminals. As ion concentration changes, a post-synaptic potential is generated. This postsynaptic potential change can either be inhibitory or excitatory, for decrease and increase of voltage, respectively. Neurons integrate synaptic inputs both spatially and temporally, so the postsynaptic output can vary markedly from the input.

2.1.2 Electroencephalography

Activity of single neurons in the human brain cannot be recorded without opening the skull. Non-invasive recordings of brain function can be obtained safely from healthy volunteers. EEG is a continuous mass-action level measure of electric potentials associated with neuronal activity. The first human recordings of EEG were conducted by Hans Berger and colleagues in the 1920s (Berger, 1929). EEG offers a millisecond-level measure of the human brain function, but it is restricted in spatial accuracy compared for instance to MEG due to volume conductor effects: the electric brain-generated signal is distorted in many ways due to skull and skin physiology when it has to be measured from the scalp (Hämäläinen et al., 1993).

What, then, are the origins of a continuous EEG signal, electroencephalogram? The most important contribution to EEG (and MEG) signals is from the pyramidal cells (Hämäläinen et al., 1993), which form majority of the neurons in the human cerebral cortex (Nieuwenhuys, 1994). These neurons have long axons and at least one of their spiny dendrites aligned perpendicular to the cortical surface, and may additionally have long horizontal axon branches allowing intralayer communication. When neurons are activated in synchrony the postsynaptic potentials they generate summate up within a few milliseconds. This is because postsynaptic potential changes are slower than for instance action potentials that last only a 1–2 milliseconds (Kandel et al., 2000). The spatially more organized structure of pyramidal cells compared to nonpyramidal ones, with evenly placed synapses, allows spatial summation of both excitatory and inhibitory postsynaptic potentials. Activation of thousands of neurons generate potential high enough to be detectable with the current EEG/MEG equipment.

The potential summation finally leads to an extracellular current flow, *primary current*, which in turn results in *return* or *volume current* when the ions are under an electric field in a conducting medium. EEG and MEG measure the electric potential distribution and the electromagnetic field, respectively, of both primary and volume current (Hämäläinen and Hari, 2002), but the ultimate goal is to model the primary current. The large size and parallel orientation of neighbouring pyramidal cell dendrites increase the combined electromagnetic field they generate compared to nonpyramidal cells (Westbrook, 2000). The distance from a pyramical cell soma to a cluster of terminals in dendrites can be for instance 500 µm (total dendritic length of cell $\sim$10 mm), whereas nonpyramical cells extend maybe 100–200 µm (total length $\sim$3 mm), with dendrites projecting to several different directions from the soma (Nieuwenhuys, 1994).

EEG is measured using electrodes placed on the skull. EEG electrodes are typically well-conductive metallic rings (in PI, silver/silver chloride;

Ag/AgCl), where the conductivity is further enhanced by electrode paste. As EEG voltage is always a potential *difference* between two points, at least two electrodes are needed. In PI, we used an electrode cap, where the 32 electrodes were evenly placed on an elastic cap, with one designated reference electrode. More dense EEG electrode configurations with for instance 128 or 256 electrodes are typically used to obtain better localization accuracy for the source signal. EEG signal, however, may not provide the best signal for source localization, as the signal is attenuated, blurred, and distorted by brain structures and the skull (see Hämäläinen et al., 1993). EEG signal is very small in amplitude, so it requires amplification before further analysis and storage. EEG systems have had for decades built-in analog-to-digital converters to allow digital data storage for further analysis. Modern EEG setups use an active electrode system, where the source signal is amplified by a small circuit placed on the electrode, but the data in PI were collected with an EEG setup using a designated amplifier.

2.1.3 Magnetoencephalography

The main source of EEG signal, extracellular current flow, is electromagnetic, so electrical activity is only one viewpoint for this signal because a magnetic field is generated as well. A continuous measure of this magnetic field is called magnetoencephalogram (MEG), which is used in PII–PIV of this Thesis. MEG is a newer imaging methodology than EEG, and was first recorded in 1968 by David Cohen (Cohen, 1968).

MEG signals are very weak compared to for instance the constant magnetic field of the earth, so MEG recordings require very sensitive sensors and preferably a shielded room to reduce the effect of external magnetic fields (Hämäläinen et al., 1993; Hari et al., 2000). The current MEG machinery use a set of superconducting quantum interference device (SQUID) sensors, located in liquid helium in a head-shaped array (Hämäläinen et al., 1993).

MEG signals can be analyzed in sensor space (like was done in PIII), similar to what is often used in EEG studies when statistical analyses are done to the signals from separate electrode channels. The more advanced option for both EEG and MEG is to use source modeling, for instance by assuming that the measured brain activity at some time point is generated by a point-like equivalent current dipole (ECD; Hämäläinen et al., 1993). This assumption is reasonable for the early evoked fields generated by a simple stimulus, as in PII–PIV, and can typically explain over 90% of the magnetic field at around the peak latency. In PII–PIV the sounds were presented binaurally, so this calls for two dipoles, one per each hemisphere. In these studies, the dipoles at around 100 ms peak from the sound onset were at posterior supratemporal plane, location typical to auditory studies (e.g., Hari, 1990; Pantev et al., 1995). However, higher-level functions or more complex stimuli require more complex models as well, such as multi-dipole models. At least in these cases, a single ECD then models the center of gravity of the stimulus-related activity, not necessarily the activity the real, physiologically plausible locations anymore. An alternative to dipoles is to use distributed source models, such as cortically constrained minimum norm estimates (MNE; Dale et al., 1999; Fischl et al., 1999) to model the brain activity and to complement the dipole results, as was done in PII.

EEG and MEG signals arise from the same neural source, but the results from studies using either EEG or MEG should be compared to each other with caution. While EEG shows signals generated basically at any point at the cortical surface, regardless the orientation, MEG is more sensitive to superficial and tangential sources that are located in fissures perpendicular to the scalp. In this context, the source orientation refers to the ECD orientation modeling the primary current, the localized brain activity, and is roughly the same as how the pyramidal cells are physiologically aligned. Compared to EEG, MEG sensors cannot detect source signals that are generated at cortical sites with exactly radial surface normal (Hämäläinen et al., 1993), often referred to as radial sources. In practice, however, the "blindness" of MEG to radial sources is not a big issue due to complex cortical folding and non-concentric conductor boundaries, and further only 5% of the cortical surface is nearly radial (0–15°; Hillebrand and Barnes, 2002). The MEG sensitivity to cortex shape shows up also as a surface bias in the estimates. As a result, only deep sources and few-millimeter grooves at the crests of gyri show very weak or non-existent MEG signals and are better detected by EEG (Hillebrand and Barnes, 2002; Goldenholz et al., 2009). The differential sensitivity to source depth and to tangential and radial sources that in part cancel each other could explain why sometimes the results from EEG and MEG, obtained either simultaneously or from different sessions, differ from each other (Raij et al., 1997; Ahlfors et al., 2010b).

2.1.4 Evoked responses

A continuous measure of EEG or MEG is nearly impossible to interpret to the extend needed in many neuroscientific studies. Only distinct abnormalities for instance during sleep and activity generated outside the brain, such as eyeblinks and muscle artifacts, are visible in the raw signal. Therefore, an averaging method is often used to increase the signal-to-noise ratio, even though the current trend is at least try to use the single-trial data whenever possible (see, e.g., Delorme and Makeig, 2004; Debener et al., 2006). To obtain an event-related response, say to a flash of light or a sound, a period of time following the repeating stimulus is recorded and averaged. This reduces the background neural activity not time-locked to the stimulus and reveals the neural activity that repeatedly follows an event. The background "noise" is a reflection of both ongoing brain activity and real noise (e.g., noise from electronic components and physical movements). The evoked response typically has fluctuations that can be categorized and compared further with different task conditions. This evoked response is often called as an event-related potential (ERP) in EEG, and event-related field (ERF) in MEG.

We utilize in all PI–PIV the most prominent response to sounds played in isolation, auditory N100 response (for EEG), or its magnetic counterpart, N100m (for MEG, where 'm' stands for 'magnetic'). The 'N' in the name refers to 'negative' voltage in EEG, which is true if the reference electrode is suitably selected, for instance EEG data recorded from channel Cz, at the vertex (midline center, top of the scalp), that is referenced to the nose. The ending '100' tells when the evoked response typically peaks around 100 ms after sound onset. This latency, however, is not fixed, but changes as a function of basic sound features like intensity and dominating frequency (Stufflebeam et al., 1998; Roberts et al., 2000). Additionally, latency variation is very strong when the sound is masked by a simultaneous masker sound (for instance, Hari and Mäkelä, 1988; Okamoto et al., 2007). N100 amplitude is very selective to the sound characteristics, and it has both frequency specific and nonspecific components (Näätänen and Picton, 1987). In fact, the frequency specificity of the auditory N100 response, a feature exploited in PI and PII, was shown already by Butler (1968), who found that the N100 amplitude to a 1000-Hz test tone increases when increasing the frequency difference between a preceding tone and the test tone.

2.2 The human auditory system

The human auditory system is responsible for transforming the mechanical energy from pressure changes in a medium (typically air), or sound, to neural activity. This process has to be efficient due to the transient nature of the sound. For instance, if we do not hear somebody speaking properly, we have to ask them to repeat what they were saying, we cannot “go back” and study the sound in more detail after hearing it. This is in contrast to vision, where when having troubles seeing something we can simply look closer in order to focus on the important aspects, assuming the image does not vanish, with top-down influence more slowly enhancing the perception. Auditory system, however, preferably has to be pretuned to listen to fine details of the sound. If this system is not in the speech-comprehension ‘mode’, for instance, naive subjects cannot tell the meaning of a specially constructed speech sound where the formant transitions are replaced by sine-wave sweeps (Remez et al., 1981), they will only hear the strange-sounding noise. Only after training or brief exposure to the original speech sound the subjects will hear the sine-wave speech as speech and can decrypt the message.

2.2.1 The periphery: auditory pathway before the cortex

Human ear is divided to three functional parts: outer, middle, and inner ear. Even the external parts of the ears, the pinnae, have a function in hearing as they amplify certain frequencies in order to help to localize where the sounds are coming from (e.g., Roffler and Butler, 1968). The sound from the outside environment is transferred through the auditory canal to the middle ear, bounded by the tympanic membrane. The pressure changes then set the eardrum (also known as tympanic membrane) in motion.

Middle ear is a small cavity which holds the three small bones (malleus, incus, stapes) that are linked together, together called the ossicles. They mechanically change the vibrations to be more suitable for the inner ear, cochlea, where the sound travels in dense cochlear fluid. Further, middle ear has muscles to stiffen the ossicles in order to protect the inner ear from very loud sounds, and to suppress incoming sound during own speech production (Borg and Zakrisson, 1975).

Cochlea is the main structure that defines the inner ear and it is the structure where the sound is finally transformed to neural signal. Sound travels as a traveling wave by displacing the basilar membrane (BM), where the inner and outer hair cells (IHCs and OHCs) are located. A single human cochlea has about 12 000 OHCs and 3 500 IHCs (Dallos, 1992), resulting a total of over 30 000 hair cells in both ears. The OHCs contribute actively in hearing, increasing the dynamic range and enhancing the frequency selectivity of the hearing (Dallos, 1992). The BM is coiled up in a spiral form, narrower at the beginning, and wider and stiffer at the end which is physically at the center of the spiral. This anatomical feature results in a gradual frequency progression, high-frequency sound maximally displacing the BM and the hair cells attached to it at the beginning, and low-frequency sounds at the end. This is called tonotopy (or cochleotopy), and this basic organization principle is maintained all the way to the cortex (e.g., Formisano et al., 2003; Talavage et al., 2004).

The auditory signal is transmitted from the IHCs through the auditory nerve to several lower nuclei before the cortex. Besides the relatively straightforward place code described above, where each frequency corresponds to a place in the BM, the neural firing in the auditory nerve follows the envelope of a sound (Cariani and Delgutte, 1996a,b) especially for complex sounds (Cedolin and Delgutte, 2005). This allows the pitch of a complex sound to be encoded in the neural firing patterns even for sounds lacking the fundamental frequency. Pitch is defined as the relative frequency sensation of the listener and it is typically close to the actual frequency content in the case of simple tonal sounds.

After the auditory nerve, the auditory feedforward system first synapses in the brainstem, in cochlear nucleus. From there, sounds travel via lateral lemniscus to the inferior colliculus (IC), either directly or synapsing at the superior olivary nucleus. The final synapse before the cortex is right above the brainstem, in medial geniculate nucleus (MGN) of thalamus.

2.2.2 Auditory cortex

Auditory cortex in primates occupies a marked portion of the cortical surface if it is defined as the cortical area responding to sounds: majority of superior temporal lobe but also large portions of prefrontal and parietal cortex can be activated by sound, task-dependently (Poremba et al., 2003). This, if course, does not tell that the areas mapped as responsive to sounds do not have other function, quite the contrary. Differentially specialized areas help integrate the sensory information, and as neurons even in the primary sensory areas responding to a single modality seem to be the minority, recent studies have suggested a larger functional role of for instance the auditory cortex (Ghazanfar and Schroeder, 2006; Winer and Lee, 2007).

The human auditory cortex is comparable in complexity to non-human primates (Kaas and Hackett, 2000). One widely used model is that human AC is organized hierarchically (Wessinger et al., 2001), similar to other primates (Rauschecker et al., 1995; Rauschecker and Tian, 2000). The primary area or field (referred to as A1) where the sensory information first arrives is most responsive to simple tones. The primary area is surrounded by non-primary, secondary areas responding preferentially to complex sounds such as band-passed noise. There is empirical and theoretical work suggesting that the sounds project from primary field to anterior "what" and posterior "where" streams, for sound feature and spatial location analysis, respectively (Rauschecker and Tian, 2000; Ahveninen et al., 2006). It should be noted, however, that this functional organization is not the whole story, as some neurons in primary areas also have preference to wider bandwidths (Recanzone et al., 2000), and even show higher-order specificity (see Nelken, 2008). Species-specific sounds, such as animal calls or speech in humans, are mostly processed in the secondary areas (for a review, Rauschecker and Scott, 2009). Processing of these more complex sounds is done via the aid of multisensory integration sites where auditory and visual information converge, such as superior temporal sulcus (STS; Poremba et al., 2003; van Atteveldt et al., 2004; Beauchamp et al., 2004).

As mentioned before, tonotopic organization starts from the cochlea, where the mapping is a simple monotopic progression of frequency. Tonotopy in human AC, however, is a more complex matter, as there seems to be several tonotopically organized mirror-symmetric fields in both primary and secondary areas (Kaas and Hackett, 2000; Lütkenhöner et al., 2003; Talavage et al., 2004; Woods et al., 2009; Humphries et al., 2010). This results in differential evoked response source locations for high and low frequency sounds (Romani et al., 1982; Pantev et al., 1995). Interestingly, the tonotopic and sound periodicity progression seem to be orthogonal, as measured by evoked MEG signal sources (Langner et al., 1997).

The organization of AC supports its role in active perception, its function can rapidly reorganize based on task requirements (Scheich et al., 2007; Jääskeläinen et al., 2007). The sound representation in the primary AC has been shown to closely follow the perception in monkeys (Petkov et al., 2007), and human studies show strong correspondance with auditory cortex function and behavior at least close to the threshold (see, e.g., Parasuraman and Beatty, 1980). Some studies show that human primary auditory cortex is more truthful to the sensory input and is not modulated by top-down effects such as attention, whereas secondary areas show stronger modulation (Petkov et al., 2004; Okamoto et al., 2007). However, a recent intracortical study shows attentional modulation in both primary and secondary auditory areas (Bidet-Caulet et al., 2007).

2.3 Attention system in the brain

Attention is an important concept for normal brain functioning. We all know how our senses sensitize when we closely attend something around us, and especially what can happen when we do not pay attention. Psychological models of attention have divided the attentional systems to several subparts. One of the basic concepts is how we voluntary choose one channel or object to further processing.

The neural basis of selective attention has been studied a lot in the visual system. There, the executive attentional control network seems to have two major subsystems: 1) dorsal frontoparietal network and 2) right ventral frontoparietal network (reviewed in Corbetta and Shulman, 2002). The dorsal network, including intraparietal and superior frontal cortices, is responsible for both preparation and maintainance of attention to goal-relevant stimuli, mostly in top-down manner. The ventral network, including right temporoparietal and right inferior frontal cortices, is activated by involuntary shifts of attention due to sudden, unexpected events, or bottom-up processing. These same attentional control systems seems to hold for the auditory attention as well (Shomstein and Yantis, 2004), although there are some differences between them (Salmi et al., 2007). Still, several auditory studies have shown the involvement of prefrontal and parietal sites (e.g., Tzourio et al., 1997; Alho et al., 1999; Zatorre et al., 1999; Shomstein and Yantis, 2006; Degerman et al., 2006; Salmi et al., 2009).

2.4 Multisensory processing

Perceiving the world around is naturally multisensory. We integrate information from different senses to form a rich model of the environment, and to increase efficancy in the typical real-world case of noisy sensory input. Understanding speech, for instance, is easier if our brains generate internal representations that integrate auditory, visual and motor information from the senses and our previous knowledge of the speech structure. Recently, it has been proposed that the whole organizational principle of the cortex is multisensory (Ghazanfar and Schroeder, 2006), and that investigating a single modality at a time gives a limited view of the function. Large number of multisensory connections are typical for normal functioning of the brain, but multisensory interactions and reorganization of sensory areas are even more pronounced in sensory-deprived humans such as blind and deaf (Kujala et al., 2000).

The pioneering neurophysiological works in the area of multisensory integration were done with single-cell recordings from superior colliculus (SC) of a cat (reviewed in Stein and Meredith, 1993). Auditory, visual, and somatosensory information converge in SC, and single SC neurons can show complicated firing patterns based on the relative strenghts of the inputs and their temporal and spatial proximity. SC is an anatomically small structure that lies deep in the midbrain, below thalamus, so human data showing specific multisensory effects in SC or other early structures are scarce (see, e.g., Musacchia et al., 2006, 2007) and the section below handles mostly human data showing cortical effects.

2.4.1 Audiovisual integration of information

Audiovisual interactions are a special case of multisensory processing, partly because of the special nature of speech in our communication, but also because both allow inflow of information from a distance, from an external event. The visual system is connected to the auditory cortex both directly (Falchier et al., 2002; Rockland and Ojima, 2003; Cappe and Barone, 2005) and through heteromodal areas (Lewis and Van Essen, 2000; Schroeder and Foxe, 2002; Kayser and Logothetis, 2009), and through subcortical thalamic projections (Cappe et al., 2009).

The McGurk effect (McGurk and MacDonald, 1976), where the subjects perceive /da/ when auditory version of /ba/ and mouth movements from

/ga/ are simultaneously presented, is an example of multisensory nature of audiovisual speech perception. The timing of the inputs is crucial, however. For speech comprehension, this temporal window of integration is relative large, several hundred milliseconds (Massaro et al., 1996; van Wassenhove et al., 2007). Still, even faster integration of auditory and visual information can occur within milliseconds, as shown by an illusion where nearly-simultaneous auditory clicks can alter the number of perceived flashes (Shams et al., 2000).

2.4.2 Speech

Speech is processed in the brain differently from sounds sharing similar complexity (e.g., Belin et al., 2000; Leaver and Rauschecker, 2010). This is because understanding and parsing speech requires an interplay between a number of high-level cortical areas.

A few alternative theories tackle on how speech is fundamentally processed in the brain. The basic idea behind the popular motor theory (or theories) of speech and the more recent mirror neuron theory (Liberman et al., 1967; Liberman and Mattingly, 1985; Rizzolatti and Arbib, 1998; Liberman and Whalen, 2000) is that the speech comprehension occurs via transforming the perceived speech to the intended or actual (see Diehl et al., 2004) articulations of the speaker. These articulations or small gestures such as lip and jaw movements are the objects on how speech is spliced up in the brain, and by grouping these basic invariant objects together the syllable and word comprehension are ultimately possible.

The alternative approach to speech perception is that speech is like any other sound we hear and we analyze the acoustical features of it from the bottom up, from simple to complex. What seems the most probable explanation of speech perception is that motor system is indeed strongly involved, but is not necessary for it (see, e.g., Watkins et al., 2003; Sams et al., 2005), or that the speech motor and acoustic processing occurs in a complementary fashion within the posterior and anterior auditory processing streams, respectively (Jääskeläinen, 2010).

2.5 Top-down effects in the auditory system

The human auditory system function can be modified during different task requirements. This modulation is especially strong in the cortex, where higher-level operations are done, but can be seen at earlier and lower levels of the system. For instance, selective attention enhances and modifies the response patterns in the AC (e.g., Hillyard et al., 1973; Hari et al., 1989; Rif et al., 1991), but modulation can be seen at lower levels such as in IC (Rinne et al., 2008) and even in cochlear activity (Giard et al., 1994; Maison et al., 2001).

Besides attention, for instance memory and speech perception can have profound influence on the auditory-cortex responses. Auditory cortex activation can even occur without sound, during auditory imagery (e.g., Halpern and Zatorre, 1999; Schürmann et al., 2002; Jäncke and Shah, 2004; Bunzeck et al., 2005), during omission of sound while listening to a piece of familiar music (Kraemer et al., 2005) or a regular pattern of sounds (Raij et al., 1997), and during a lipreading task without sound input (Calvert et al., 1997; MacSweeney et al., 2000; Pekkola et al., 2005).

The auditory cortex has several functional organization principles, such as the tonotopy introduced in the last section. Due to these functional maps and multitude of local connections, it has recently been proposed that the auditory cortex always adapts to task needs depending on the goal (Scheich et al., 2007; Jääskeläinen et al., 2007). The benefit of gain or selectivity increase during task conditions is to enhance for instance stimulus detection or discrimination, both task types actually relying on overlapping but dissociable brain networks (see Tramo et al., 2005).

2.5.1 Effects of selective attention

Attending closely a train of standard tones allocates resources in the brain when we continually try to match the new sound to the internal template. Typical finding is that during a focused attention task the sensory-specific responses are modulated compared to a passive baseline. One view to the attention-modulated brain activity is the direct enhancement of neuronal assemblies generating the N100/N100m response. Attentional gain control (Hillyard et al., 1998; Kerlin et al., 2010) has been an influental theory for several years, showing that attention increases the stimulus-evoked brain activity by some magnitude. Studies in the visual system have shown that whether this gain increase occurs on its own, or whether there are selectivity increase may depend on the stimulus characteristics and the relative size of attentional foci (Herrmann et al., 2010), or other objects in the visual field (Lee and Maunsell, 2010). One solution to this confound is convergence to a model where both gain and selectivity increases may take place simultaneously (e.g., Reynolds and Heeger, 2009; Lee and Maunsell, 2009). This kind of weighted model has recently been proposed also for multisensory influence (Ohshiro et al., 2011).

In auditory studies, there are numerous reports of increase in neural response amplitude, i.e., gain control, when attending to relevant sounds. This is the case in several research modalities: in EEG (Hillyard et al., 1973; Woldorff and Hillyard, 1991; Tiitinen et al., 1993; Alcaini et al., 1994), MEG (Hari et al., 1989; Rif et al., 1991; Woldorff et al., 1993; Fujiwara et al., 1998), functional magnetic resonance imaging (fMRI; Pugh et al., 1996; Grady et al., 1997; Jäncke et al., 1999; Petkov et al., 2004; Rinne et al., 2005; Woods et al., 2009), and in positron emission topography (PET; Alho et al., 1999; Hugdahl et al., 1999; Zatorre et al., 1999; Alho et al., 2003).

The alternative view to the observed response modulation during selective attention is that it manifests from attention-related activity independent of stimulus-evoked, exogenous N100/N100m generation. This evoked response component has been named processing negativity (PN) (see Näätänen, 1982, 1992), or processing field (PF) in MEG (Hari et al., 1989). PN/PF is calculated as the negative difference wave (Nd/Ndm) by subtracting attended from non-attended ERPs/ERFs. This negativity starts from about 50–100 ms after the stimulus, peaks highest at 180–200 ms, and lasts for several hundred milliseconds (Näätänen, 1982). Nd features are sensitive to the experimental paradigm. It has a complex task difficulty dependency so that a more difficult task with higher sound presentation rate of sounds often reduces the latency of Nd compared to an easy task (see Näätänen et al., 1992). However, setting the experiment more demanding by reducing the sound discriminability increases the latency and reduces the amplitude of Nd (e.g., Hansen and Hillyard, 1980; Alho et al., 1987; Nahum et al., 2009), sometimes with amplitude increase especially with subjects with poorer performance (Nahum et al., 2009).

Due to methodological limitations EEG or MEG cannot accurately answer for a single experimental dataset which attention-related enhancement model is better suited because separating both spatially and temporally overlapping signals is practically impossible. For instance, besides Nd, auditory N100/N100m response is known to be sensitive to the presentation rate (Hari et al., 1982; Näätänen and Picton, 1987) and has several spatially distinct subcomponents with different properties (Näätänen and Picton, 1987). Thus, auditory attention studies with similar paradigms and similar attentional enhancements can be interpreted using either model: either by specific modulation of N100/N100m subcomponents or by summation of task and feature-specific attentional processes, with N100/N100m generators mostly intact. There are, however, evidence from intracortical recordings in humans that support the notion that N100/N100m and later-latency component P200 are directly enhanced during attention (Neelon et al., 2006; Bidet-Caulet et al., 2007; Neelon et al., 2011) instead of supplemental attention-induced activity explaining the response enhancement. But, notably even these studies cannot definitely rule this PN/PF view out due to very focal recording sites. Still, explaining the experimental data by extra processes and components that are very sensitive to the sound presentation rate (Neelon et al., 2011) and task demands might lead to an overly complex model of brain function during active perception. This is especially the case when the same data could be explained in a neurobiologically plausible way such as by differentially adapted neuronal populations instead of distinct task-specific processes (see, e.g., Ulanovsky et al., 2003; Jääskeläinen et al., 2007; May and Tiitinen, 2010).

2.5.2 Effects of lipreading and own speech production

During audiovisual speech perception, lipreading affects early auditory processing at ∼100 ms from sound onset, by enhancing (Giard and Peronnet, 1999; Hertrich et al., 2007) or suppressing the N100/N100m amplitude (Klucharev et al., 2003; Besle et al., 2004; Jääskeläinen et al., 2004b; van Wassenhove et al., 2005). The stimulus congruency, whether the auditory and visual stimuli are compatible, affects the responses (Sams et al., 1991). The lipreading-induced suppression involves simple tone-evoked N100m amplitude as well (Numminen and Curio, 1999; Curio et al., 2000; Houde et al., 2002), similar to the suppression effect during own speech production (Numminen et al., 1999). Even earlier effects at the level of brainstem during a lipreading task have been shown (Musacchia et al., 2006). And, in fact, producing speech suppresses the auditory system at low frequencies already before the cochlea by stiffening the middle-ear stapedius muscle (Borg and Zakrisson, 1975).

2.5.3 Neural substrate for top-down effects

Animal models have shown that cholinergic modulation from the nucleus basalis can rapidly reorganize primary auditory cortex function and receptive fields (RFs; Weinberger, 1998, 2004), and that the amount of acetylcholine (ACh) during encoding affects the specificity of the auditory sensory memory (Weinberger et al., 2006). Release of ACh into cortex can change the RFs for instance towards the attended stimulus frequency, with differential time course for inhibitory and excitatory synapses, resulting a cortical memory trace similar to what happens in perceptual learning (Weinberger, 1998; Froemke et al., 2007; Weinberger, 2007). This is similar to plasticity that has been observed in other sensory systems (Rauschecker, 1999) and enables reorganization of cortical networks to task requirements (Jääskeläinen et al., 2007).

Besides cholinergic modulation, noradrenergic modulation from locus coeruleus can selectively modify the neural activity, enhancing the signal-to-noise ratio of relevant signals, compatible with perceptual benefits of selective attention (Hirata et al., 2006). Noradregeneric modulation could also occur indirectly via frontal areas (see Sarter et al., 2001).

These two modulators probably serve a different purpose, as noradrenergic modulation has different dynamics than cholinergic modulation during a visual attention task (Dalley et al., 2001). Further, these two different modulators probably target different attentional networks: goal-directed top-down network is affected by cholinergic drugs, whereas stimulus-driven bottom-up network by noradrenergic drugs (see Corbetta and Shulman, 2002). Due to their slow time course, cholinergic inputs play a role especially in difficult tasks requiring sustained selective attention (for a review, Sarter et al., 2001).

3. Objectives of the studies

The general aim of this Doctoral Thesis is to investigate mechanisms of top-down influences in the human auditory cortex of healthy volunteers by means of non-invasive electrophysiological recordings (EEG and MEG). The studies focused on the effects during three task conditions: 1) under selective attention, when responding to sound features (PI and PII), 2) while lipreading (PIII and PIV), and finally, 3) during a silent speech-production task (PIV). The basic idea is to investigate what happens to the auditory cortex when we are focusing our attention willingly *towards* the sounds compared to baseline, or when we are focusing our attention *away* from the sounds by doing a visual task that either does have some (PIII) or does not have any clear relevance (PI, PII, and PIV) to the presented sounds. Previous literature suggests that selective attention enhances the neural responses, most typically by increasing the neural gain (Hillyard et al., 1998), that is, by increasing the response amplitudes. The crossmodal effect during a lipreading task (Klucharev et al., 2003; Besle et al., 2004) and speaking either silently or aloud (Numminen and Curio, 1999) is the opposite, auditory-evoked neural responses to external sounds are suppressed in amplitude.

The specific objectives of the individual studies included in the Thesis are listed below, main study question(s) *emphasized*.

In PI, we investigated how attention influences the population-level frequency tuning. We measured the N100 amplitude with EEG as a function of notch width using continuous notched-noise maskers and tones presented at the center of the notch. Previous study dating years back (Sams and Salmelin, 1994) estimates the frequency tuning in passive-only condition. Here, our study question was *does the gain model hold, or does attention increase the feature selectivity as well during an active listening task*. We specifically hypothesized here that gain-only mechanism would increase the response with all the notches, whereas frequency selectivity increase would modulate the amplitudes differentially with narrow and wide notches.

In PII the attentional enhancements were explored in more detail to complement the previous results. With MEG we asked 1) *where the attentional modulation occurs* by localizing the responses to auditory areas, 2) *are there are hemispheric differences or temporal dynamics in the attentional modulation*. Here, we used longer auditory stimulus than in PI in order to see the potentially differential changes in the initial sound onset-sensitive N100m response and in the later-latency sustained response.

PIII used MEG to explore specificity of the lipreading-related auditory-cortex suppression effect. We used frequency-modulated (FM) tones sweeping from low to high frequencies, or vice versa, as our auditory stimuli. The FM sweeps are simplified versions of the formant transitions used in speech. Our main goal here was to study *whether the N100m response modulation during lipreading is specific to the presented sound*. We hypothesized that lipreading /ga/ will have different effect on the ascending and descending frequency slopes than lipreading /ba/, because the /ba/ and /ga/ sounds have ascending and descending second-formant transitions, respectively.

In PIV, we hypothesized that lipreading task modulates early auditory processing, and this processing can be seen even at the level of simple tone-evoked MEG responses. We asked 1) *whether lipreading has differential effect on the frequency range critical to speech perception* (approximately 500–4000 Hz), and 2) *is the suppression effect both for the lipreading and covert speech production tasks quantitatively and qualitatively similar*. We envisioned that possible similarity during the two tasks would suggest whether the suppression effect shares the same neural basis.

4. Summary of Publications

This chapter summarizes the main methods and findings of PI–PIV. The aim of this chapter is to give an overview of the studies, as the actual studies hold the fine-grained details necessary for replication or a follow-up study. Since the used methodology and means of data collection and analysis were quite identical in all individual studies reported in PI–PIV, only the stimuli, results and implications from each study are handled separately. Table 4.1 briefly summarizes the methodological similarities and differences of the studies.

4.1 Materials and methods common to all studies

All of the experiments were carried out at Aalto University School of Science. The first one, reported in PI, applied only EEG, with 32-channel EEG system (Brain Products GmbH, Germany) that was located at the Department of Biomedical Engineering and Computational Science. The measurements were carried in a shielded room, providing attenuation for both external acoustical and radio frequency (RF) interference. MEG recordings were done at Low Temperature Laboratory, Aalto University School of Science, using a 306-channel whole-head neuromagnetometer (Vectorview, Elekta Neuromag). The MEG device was located in three-layer, magnetically shielded room, so external electromagnetic interference was minimal.

The subjects included in the studies were all healthy volunteers, mostly young university staff and students. Their eye sight was normal or corrected-to-normal and hearing ability normal, based on self report.

Stimulus delivery was computerized to allow stimulus event triggering to the EEG/MEG device. In PI, the sounds were presented using a set

Pub.	Method	Subjects	Auditory stimuli	Visual stimuli	Task of the subject
PI	EEG	N=20, age 18–28 years, mean 23	Tones (100 ms, 1000 Hz) at hearing threshold with continuous notched noise	Fixation cross; Silent movie	Oddball, detect deviants (auditory task) / Watch movie (passive control)
PII	MEG	N=14, age 21–46 years, mean 29	Tones (300 ms, 1000 Hz) at hearing threshold with continuous notched noise	Gabor patches	Oddball, detect auditory or visual deviants
PIII	MEG	N=8, age 22–32 years, mean 26	FM sweeps (50 ms, frequency range 200–2800 Hz), 55 dB over hearing threshold (clearly audible)	Visual articulation of /ba/ and /ga/ or still face	Detect visual stimulus change (block level, from one stimulus type to another)
PIV	MEG	N=19, age 20–32 years, mean 24	Tones (50 ms, 125–8000 Hz), 55 dB over hearing threshold (clearly audible)	Visual articulation of /a/, /i/, /o/, /y/; Still face; Ring expansion	Detect repetition (articulations and rings); eyes focused on mouth area (still face); covert speech production (still face)

Table 4.1. Experimental setups of the studies included in the Thesis.

of high-quality computer loudspeakers, and visual stimuli were displayed on a computer monitor. In PII, sounds were presented free-field as well, using an MEG-compatible panel loudspeaker. In PIII and PIV, acoustical stimulus delivery was done using a set of flexible tubes and ear inserts, with sound source located outside the MEG shielded room. In all MEG studies (PII–PIV), visual stimuli were identically displayed using a back-projector screen located in front of the subject.

For all of the studies, EEG and MEG raw signal was recorded with a high sampling rate (PI: 500 Hz, PII: 2000 Hz, PIII and PIV: 600 Hz) to allow more freedom in the later offline analyses. Epochs time-locked to auditory stimulus presentation were averaged to obtain ERPs and ERFs. Both ERPs/ERFs and raw data were monitored online in order to assess the data quality and to detect possible problems with subjects (excessive movements, continuous eyeblinks). For the actual analyses, offline-averaged data were used, with artifact rejection, baseline correction and lowpass filtering (typically only up to 40 Hz) suitable for the study goals.

4.2 Publication I - Selective attention to sounds increases both gain and frequency selectivity

The first study was a psychophysical experiment involving several notched-noise maskers presented together with short sinusoidal tones. The maskers were created by bandstop filtering white noise with a parametrically varying notch width. The relative level of the masker and tones were adjusted so that 1000-Hz tones were at 50% hearing threshold (Levitt, 1971) with the white noise masker. The setup was similar to an earlier study done by Sams and Salmelin (1994) with the addition of auditory attention tasks. We compared ERPs to the standard tone during both attended and unattended tasks. During the auditory task the subjects discriminated either frequency (1020 Hz vs. 1000 Hz) or duration deviants (150 ms vs. 100 ms) from the ongoing train of standard tones.

The results (Figure 4.1d) confirmed the classical finding that selective attention modulates the auditory cortex activity at around 100 ms latency by means of increased signal power (Hillyard et al., 1973). Increased neural gain alone was not sufficient to explain the response magnitude changes with different noise maskers, however, and based on computational simulations done at neural population level (Figure 4.1a,b) we proposed that enhanced frequency selectivity of auditory cortex neurons is a necessary factor in addition to increased neural gain. These results were replicated in humans using MEG (Okamoto et al., 2007), and supporting data has been obtained for instance from owl visual system (Winkowski and Knudsen, 2008).

4.3 Publication II - Frequency-specific attentional modulation was differential for the initial N100 and the later-latency sustained response

Here, we probed the auditory frequency selectivity again with tones embedded in spectrally complex notched-noise masker as in PI. This time we measured MEG signals instead of EEG in order to localize the sources accurately. Further, we included an active visual control task and added an extended selection of notches.

Using MEG, we replicated the same selective attention enhancement at the initial peak (N100m) as in PI (Figure 4.2a vs. Figure 4.1c). This

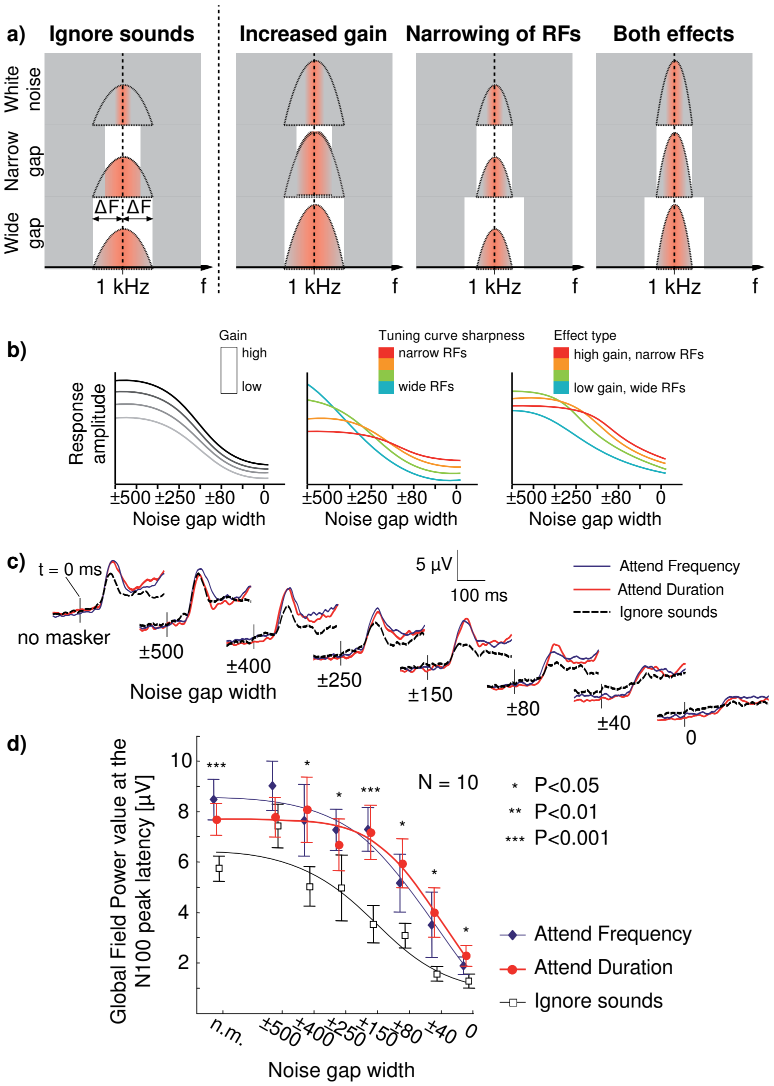


Figure 4.1. (a) Hypothesized top-down effects under selective attention in PI and PII. The bell-shaped curve in the leftmost column represents the single-neuron receptive field with 1-kHz best frequency during baseline condition of PI. There, the subjects were instructed to ignore the sounds and watch a silent movie. The noise masker, represented in grey, overlaps the single-neuron receptive field, and the red-coloured area how likely the neuron will respond. (b) Schematical results on how the various mechanisms would be seen at population level as a function of noise gap width. (c) Grand average global field power (GFP) measure of the evoked responses. (d) Results of PI that were best explained by a combination of gain increase and narrowing of receptive fields.

was, however, the case only during playback of the continous masker (Figure 4.2b). In PI, the attention effect was found both with and without masker sound, whereas in PII the attention effect was absent when the tones were played alone. The differential sensitivity of MEG and EEG to auditory cortex source orientation (Ahlfors et al., 2010a) could explain both this confounding result and why the interaction effect shown in EEG data in PI failed to reach significance for the initial N100m response recorded in MEG. The later-latency sustained response at around 300–400 ms showed interaction with the notch width (Figure 4.2c), especially in the left hemisphere. The most prominent enhancement was with the narrowest notches (±50–±200 Hz).

Our results show evidence of a dual-stage filtering mechanism of sensory input in the human auditory cortex: 1) gain increase of attended sounds at early (∼100 ms) latencies bilaterally in posterior parts of the secondary auditory areas, and 2) adaptive filtering of attended sounds from task-irrelevant background masker at longer latency (∼300 ms) in more medial regions, predominantly in the left hemisphere, enhancing processing of near-threshold sounds. This is possibly a reflection of specific adaptation in the antero-ventral “what” stream (Jääskeläinen et al., 2004a; Ahveninen et al., 2006; Rauschecker and Scott, 2009; Leaver and Rauschecker, 2010). The results of PII indicate that higher-order areas in the hierarchy show more feature-selective top-down modulatory influence than the more posterior auditory areas, where the initial and less specific auditory analysis is performed.

4.4 Publication III - Lipreading a specific phoneme results a specific suppression of auditory-evoked response

In this study, we asked whether lipreading-related auditory cortex suppression is specific to the lipread phoneme. For this, we performed an MEG experiment where we played the subjects a continuum of short frequency-modulated sine-wave sweeps while they were lipreading phonemes. The sweeps were created parametrically, from ascending to descending in frequency, to cover the second-formant transitions of /ba/ to /ga/, through /da/ (see Figure 4.3). Additionally, we played the subjects the common first formant transition shared by /ga/, /da/, and /ba/, all in random order.

The results (Figure 4.4) show that lipreading /ga/ suppressed specifically the left-hemisphere N100m response for the last formant of the continuum with the steepest descend. This is equivalent to the formant sound

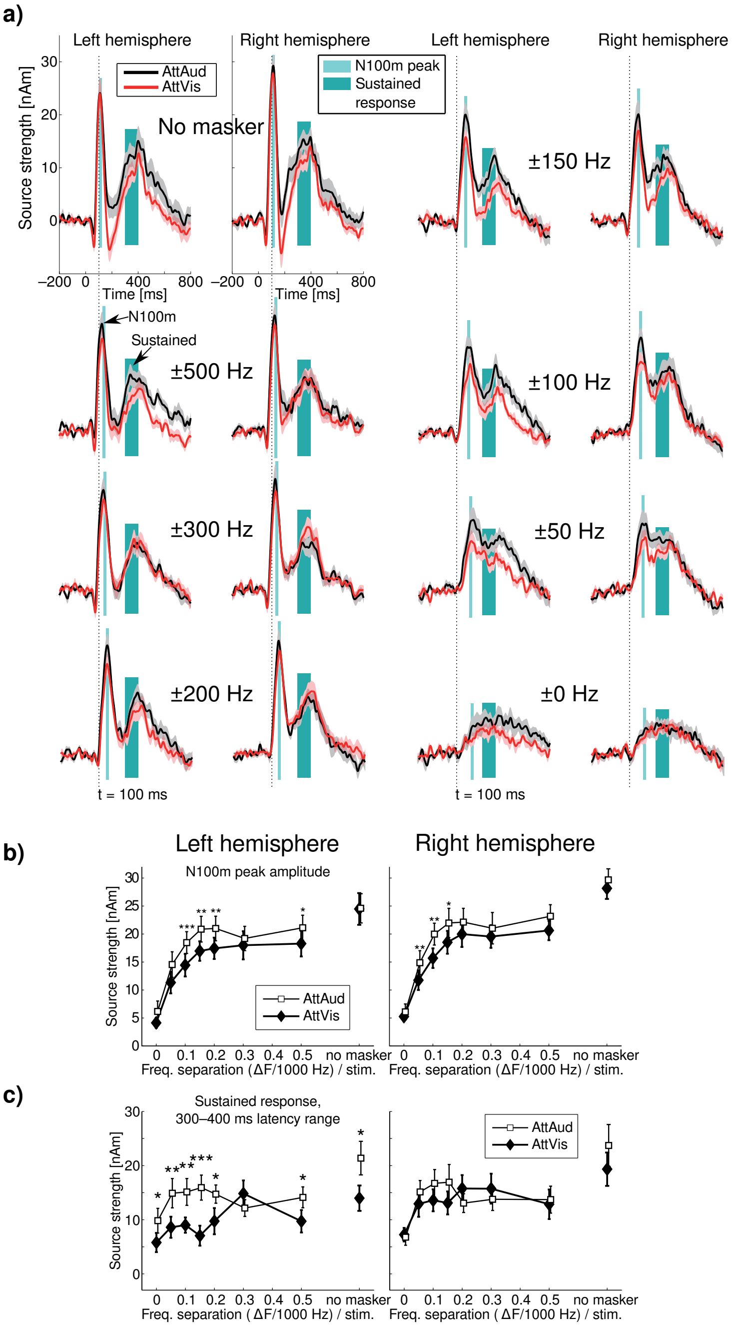


Figure 4.2. Results of PII. (a) Grand average (N = 14) source waveforms (±standard error of the mean, shaded areas) from PII for each stimulus type, hemisphere and condition. The source waveforms illustrate the gradual suppression of response amplitudes and increase in latency with narrower notches, similar to Figure 4.1c. (b) N100m source strengths were clearly modulated by masker type and attention. The attention increased the response amplitude in both hemispheres with all the notches, but this time without frequency specificity. (c) Sustained response amplitude showed frequency specific modulation in the left hemisphere with the narrowest notches.

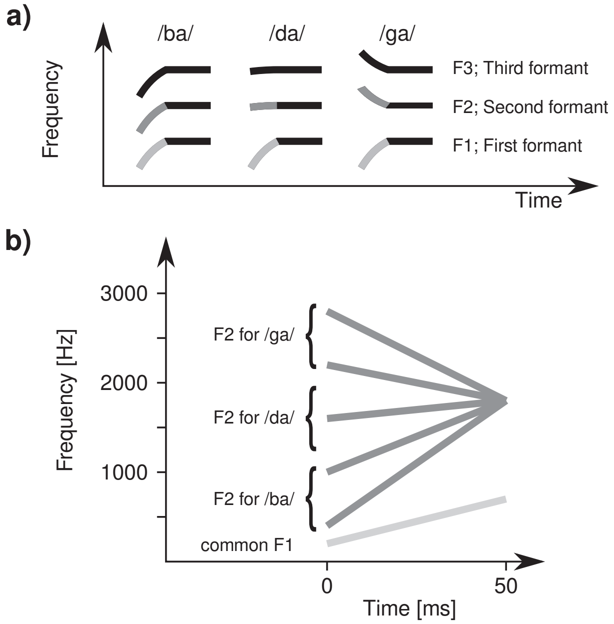


Figure 4.3. A schematic of the stimuli used in PIII. (a) Simplified time-frequency representations (spectrograms) of the syllables /ba/, /da/, and /ga/. The second-formant transition is ascending for /ba/, nearly level for /da/ and descending for /ga/. The first- and second-formant frequency transitions that were used as the basis for generating the actual stimuli are marked in shades of gray. (b) The actual FM sweeps used in the study.

that is included in the auditory /ga/. We propose that this reflects an adaptation of neuronal populations tuned to formant transitions. Additionally, we found that lipreading suppressed the first common formant in general, but more prominently while lipreading /ga/ than while lipreading /ba/.

4.5 Publication IV - Lipreading and covert speech production suppress the auditory cortex reactivity via similar mechanism

Previous studies have shown that during both observation of speech (lipreading) and producing speech the auditory cortex responses are suppressed. In this study we studied if there is frequency specificity in this suppression, and whether the effects are identical for lipreading and silent (covert) speech production. Covert speech production task here meant that the subjects were silently and without mouth movements, in a self-paced manner, producing the same Finnish vowels that they were lipreading (/a/, /i/, /o/, and /y/).

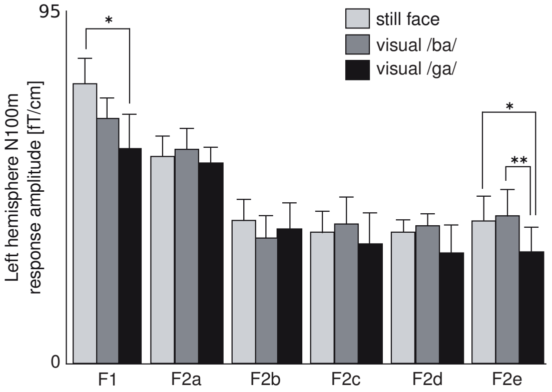


Figure 4.4. Results of PIII. Leftmost bars show response suppression for the first common formant (F1). Importantly, there was a specific suppression of the formant transition corresponding to /ga/ sound (F2e, rightmost bars) only while lipreading /ga/. © 2008 Lippincott Williams & Wilkins, Inc. Reprinted with permission.

Our results, depicted in Figure 4.5, show that during both lipreading and covert speech tasks, auditory-cortex responses to probe sounds of all frequencies were suppressed identically. In addition to the strong and generic suppression effect, a smaller suppression effect time-locked to the mouth opening during the lipreading task was observed (Figure 4.5b).

The results suggest that the top-down influence causing response suppression is possibly due to an efference copy signal from the speech production system. This means that especially in extreme cases when we are lipreading without matching sound, we are using our speech production system to help reach the goal of understanding only the lip movements.

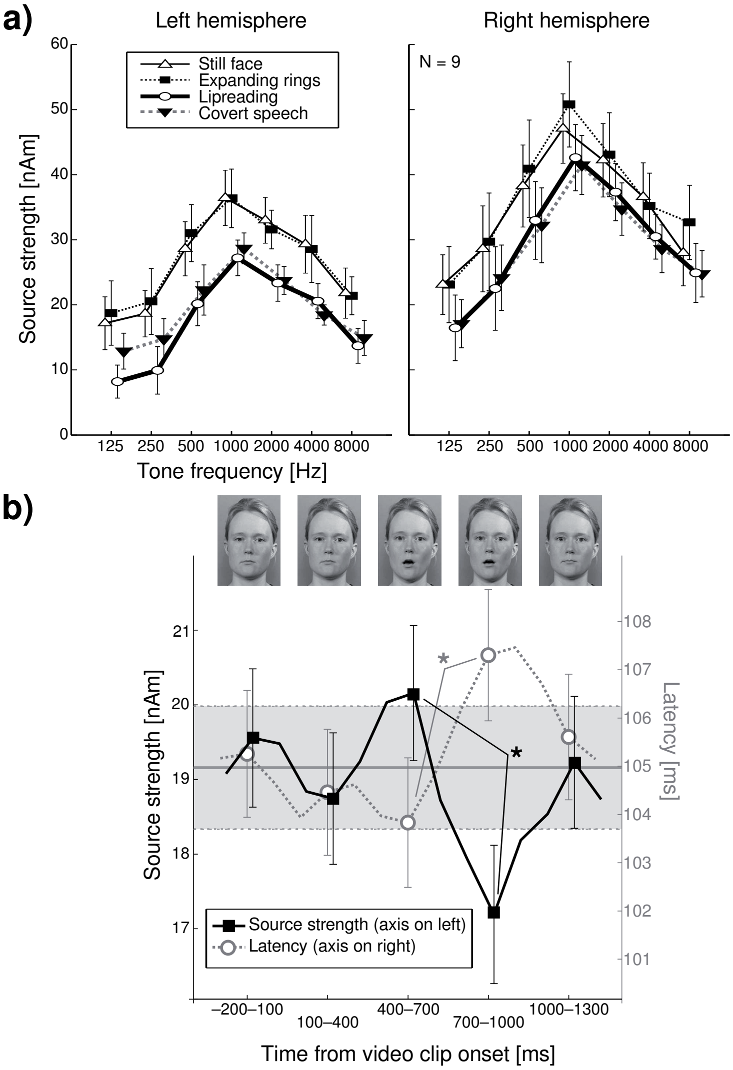


Figure 4.5. Results of PIV. (a) The response suppression effect was practically identical in amplitude at all tested sound frequencies during both lipreading and covert speech production tasks. (b) Further, during lipreading task, we observed a combination of amplitude and latency effect time-locked to mouth opening in video (from 400–700 ms to 700–1000 ms). Here, all frequencies are pooled together, and the gray area shows the mean ± standard error of the mean.

5. Discussion

5.1 Active role of the hierarchical auditory system

Taken together, the results of PI–PIV suggest that we cannot think of the human auditory cortex as a simple "black box", one input always resulting in an identical output. Indeed, human auditory cortex has been shown to be highly flexible in its function (Scheich et al., 2007; Jääskeläinen et al., 2007). Our method of investigation using a non-invasive methodology is of course not as detailed as animal single-cell studies that show strong receptive field plasticity (e.g., Fritz et al., 2003, 2005), or subdurally recorded human studies on epileptic patients that show drastic changes during attention (Bidet-Caulet et al., 2007; Neelon et al., 2011). Here, in all studies, we aimed to create a baseline to which we compared the active conditions with identical stimuli but with different task instructions. Thus, the bottom-up input reaching the ear was identical, but due to top-down information from the frontoparietal areas related to attentional control networks (Corbetta and Shulman, 2002), or visual and multisensory areas see (e.g., Arnal et al., 2009), the auditory cortical activity was strongly modified.

The auditory stimuli are processed hierarchically (Wessinger et al., 2001), with information initially flowing to primary areas and thereafter to secondary auditory association and multisensory sites. We focused on the N100/N100m response because it was the most prominent and of highest amplitude in EEG/MEG and showed the majority of attentionally and cross-modally induced changes. In PII, the later-latency sustained response at 300–400 ms from the stimulus onset was found to be more frequency specific, supporting the hierarchical organization, with several distinct processing stages (Woods and Alain, 2001; Escera et al., 2003). This is also related to an early study which shows, based on simple tone stimulus experiment, that later-latency P300 responses show the feature-specific top-down effects instantly during the course of whole experiment, compared to the N100 effect which took 30–45 s to emerge and vanished after 7 minutes (Donald and Young, 1982). In crossmodal studies, where audiovisual or speech perception is involved, in amplitude stronger later-latency response modulation is also typical besides the initial early-latency effects (Klucharev et al., 2003; Besle et al., 2004; Talsma and Woldorff, 2005), starting from 50 ms after the stimulus.

5.2 Top-down controlled enhancements and suppressions

Here we show that the results for two types of top-down influence, lipreading and intramodal selective attention, seem to be the opposite at the grand, macroscopic level of the underlying neural activity that the averaged EEG and MEG responses reflect. On one hand, selective attention enhanced the response amplitudes in both PI and PII with all of the masker sounds. Additionally, the enhancement in N100 (PI) and the sustained response (PII) amplitude was proportionally larger with narrower frequency notches. This suggests a more profound enhancement when the masker sound nearly maximally attenuates the tones. This is when more effort is required from the subject to perform in the task, and the extra boost given by selective attention is crucial in segregating the task-relevant sounds from noise. Lipreading, on the other hand, was shown to suppress the N100m responses specifically when the FM sweep, auditory “building block” of speech sound, and the visual articulation were related (PIII). In PIV, the findings of general suppression already reported before (Numminen and Curio, 1999; Curio et al., 2000) were further extended by showing the dynamics of the suppression depended on the stage of the articulation, and showing that both lipreading task and silent speech production suppressed the auditory cortex reactivity identically.

The observed enhancement in N100/N100m amplitude during selective attention in PI and PII was to be expected, and is in line with a long history of selective attention studies (Hillyard et al., 1973; Hari et al., 1989; Rif et al., 1991; Woldorff and Hillyard, 1991; Tiitinen et al., 1993; Woldorff et al., 1993; Alcaini et al., 1994; Fujiwara et al., 1998). What was novel here was the frequency specificity of the effect. The feature specificity of the auditory attention effect, in combination with gain increase, has only recently been shown besides the results of PI and PII (Okamoto et al., 2007; Neelon et al., 2011), whereas several visual attention studies in humans show both gain and selectivity increase (e.g., Murray and Wojciulik, 2004; Huang and Dobkins, 2005; Herrmann et al., 2010). What remains to be seen is the mechanism of the effect, whether sharpening of the frequency tuning is mediated for instance by thalamocortical input to layer 4 (O'Connell et al., 2011).

The alternative view to N100/N100m modulation during attention calling for independent attention-dependent neural generators (Näätänen, 1982) does not fit the present results well for instance due to complexity of the modulation with different notch widths. As seen in both PI and PII, the attention-related enhancement was not directly proportional to the task difficulty increase. Additionally, the slow presentation rate of standard sounds used both in PI and PII would predict a small attention-related enhancement, as shown in MEG (Rif et al., 1991) and fMRI (Rinne et al., 2005) studies where the rate dependency was tested. Many of the classical auditory attention studies (starting from Hillyard et al., 1973) used a dichotic listening paradigm with very high presentation rate, up to 8-fold or more compared to the one used in PI and PII, to reveal the attentional enhancements as too slow a rate could remove the effect (see Näätänen and Picton, 1987).

The novelty of these auditory attention-related enhancements is most probably related to the continuous notched-noise masker sound used to set the attended sounds very close to the hearing threshold, additionally causing an interplay between the attended tones and the masker sound (Lee and Maunsell, 2009). The continuous masker tentatively adapted the non-specific evoked response generators, letting the frequency-specific, sharply tuned neuronal populations activate in response to the attended tones. The lack of N100m attention effect in PII with no masker sound also partially confirms that the effect is specific to the experimental setup. Alternatively, this partial lack of attention effect could be interpreted so that the task difficulty should be high enough to induce attention effects with MEG, similar to what has been observed in auditory periphery (Giard et al., 1994).

In PI and PII, both dorsal and ventral attentional subsystems (see Background section and Corbetta and Shulman, 2002; Shomstein and Yantis, 2004) were likely involved due to stimulus design. The notched-noise masker sound largely suppressed the faint tones, while the subjects were working hard to discriminate the stimuli from each other. This probably resulted in a dynamic interplay with the bottom-up and top-down stimulus processing, operated by ventral and dorsal attentional networks, respectively. In fact, for audition, the bottom-up and top-down mechanisms cannot be easily separated, as they operate on highly overlapping areas (Salmi et al., 2009). Additionally, the shifts between auditory and visual tasks in PII presumably activated the dorsal subsystem, as a similar task-shifting paradigm has shown to activate posterior parietal and superior prefrontal cortices (Shomstein and Yantis, 2004). In PIII and PIV, however, probably only the dorsal attentional network was involved especially during the task changes but also during the sustained visual attention.

Simultaneous auditory and visual events can modulate both unimodal sensory cortices and multimodal areas. However, whether the modulation is enhancing or suppressing at the grand scale is not clear. Recently, a concept of multisensory integration through divisive normalization during multimodal perception was proposed (Ohshiro et al., 2011), similar to a model explaining attention effects by normalization in the visual system (Reynolds and Heeger, 2009). This normalization of the multisensory input depending on whether the inputs are optimal for the given modality can explain both super- and supra-additivity, both effect types often seen in for instance audiovisual EEG/MEG studies on speech perception (Giard and Peronnet, 1999; Klucharev et al., 2003; Besle et al., 2004; Jääskeläinen et al., 2004b; van Wassenhove et al., 2005; Hertrich et al., 2007). Functional MRI studies in humans have shown increase in both multisensory and primary sensory areas during a presentation of simultaneous audiovisual stimuli (Noesselt et al., 2007). Opposite effect types where the primary sensory areas are suppressed but multimodal areas such as STS are enhanced seem to occur when the subjects bind the audiovisual events together (Bushara et al., 2003). Alternatively, these differences could be explained by different visual modulatory pathways, a direct connection from the visual motion area to the auditory areas versus a slower feedback connection through STS that depends on the congruency of signals (Arnal et al., 2009).

5.2.1 A unified model to explain the observed modulations

All the effects could theoretically be explained by specific adaptation of the neural generators underlying the N100/N100m, as depicted in Figure

5.1. However, the present results do not directly answer to the claims, the possible adaptation mechanisms remain an open question.

5.3 Source of top-down modulatory signal

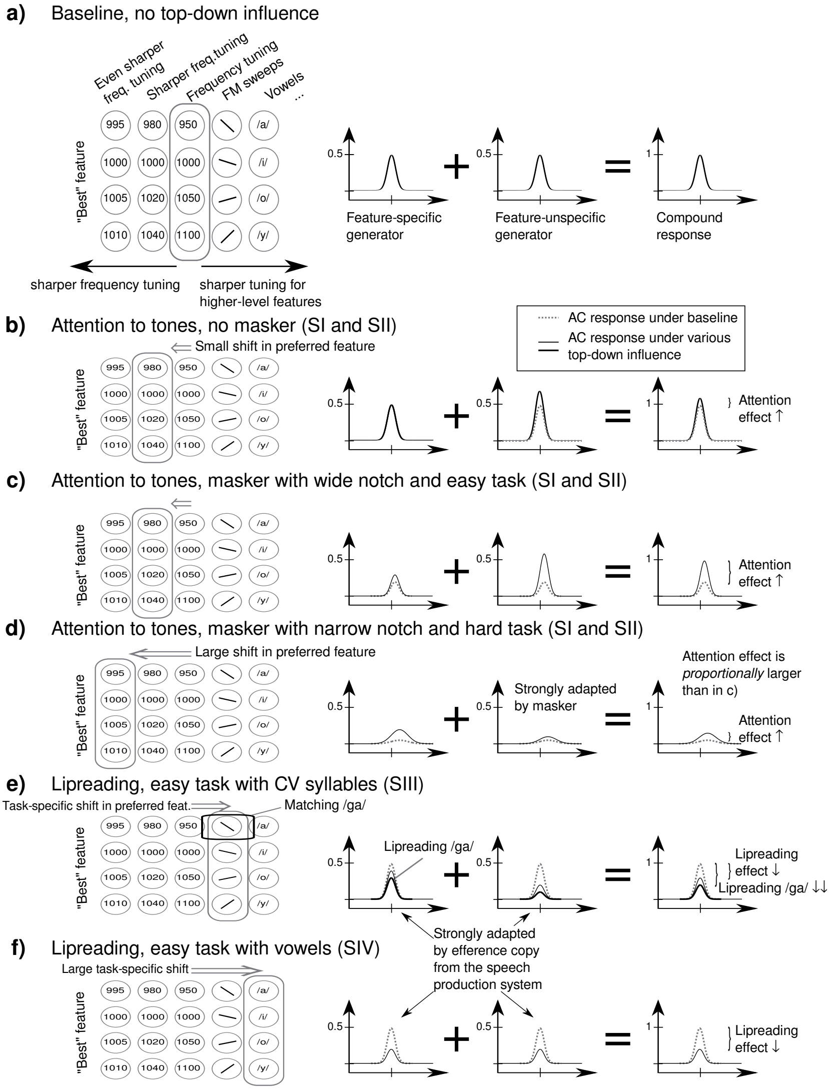
All present studies showed modulation of N100/N100m response, generated in the planum temporale. However, the top-down signal modulating the auditory-cortex activity can originate from extratemporal, non-auditory areas. In PI and PII, top-down modulatory signal under selective attention was probably from the prefrontal areas, part of the attentional control networks in the brain (Corbetta and Shulman, 2002; Shomstein and Yantis, 2004). The prefrontal areas act as a gating mechanism (reviewed in Knight et al., 1999) modulating already the middle-latency evoked responses (at around 25–35 ms) to simple click sounds (Knight et al., 1989). Patients with frontal lobe lesions show, besides poorer task performance, reduction in attention-related effects at around 80–500 ms latency in EEG (Knight et al., 1981). A recent study suggests that feedback from frontal to auditory areas is indeed crucial in conscious perception of sounds (Boly et al., 2011). Further, besides frontal areas, auditory attention tasks show involvement of posterior parietal and temporal cortices (Paus et al., 1997; Alho et al., 1999; Zatorre et al., 1999) especially during task switching (Shomstein and Yantis, 2004), so these may be considered as alternative sources for modulatory input even though they could not be distinguished in PI and PII due to methodological limitations.

In PIII and PIV, however, attentional networks underlying selective auditory attention probably did not play such a big role as the attention was directed to visual stimuli, away from the auditory stimuli. PIV shows no difference between the passive baseline and a visual control task, suggesting that a relatively easy visual attention task did not modulate auditory cortex as such. The top-down modulatory signal suppressing the MEG responses likely originated from the motor or premotor areas (see Rauschecker, 2011), not from the visual areas, as the auditory and visual stimuli of PIII and PIV were not congruent and presented as their own streams. The suppression in PIII and PIV was tentatively caused by an efference copy signal from the speech production areas during a lipreading task (Sams et al., 2005; Skipper et al., 2007). The lipreading and covert speech production tasks both activate Broca's region (Nishitani et al., 2005) in the left inferior frontal gyrus, one possible source for modulatory signal that is densely connected to the premotor and motor regions. Still, based on the present data, we cannot completely exclude the notion that the suppression was due to a direct input from the visual system via for instance the multisensory "hub", posterior STS (Schroeder and Foxe, 2002; Cappe and Barone, 2005; Kayser and Logothetis, 2009), as shown in a recent MEG study (Arnal et al., 2009). Theoretically, the suppression could be mediated by lower-level, subcortical connections from thalamus, via specific afferents from the medial pulvinar or nonspecific afferents from medial interlaminar nuclei (Jones, 1998; Cappe et al., 2009; O'Connell et al., 2011). Or, as an alternative account, the reduced amplitude could also in part be due to less but more sharply tuned neurons responding to the repeated stimuli with no behavioral relevance (see Grill-Spector et al., 2006).

5.4 Timing of the multisensory input signals

For the suppression effect during speech production, both timing of the motor command and the stimulus type are important. A recent EEG

Figure 5.1 ***(on the next page).*** (a) Hypothesized contribution from feature-specific and feature-unspecific N100/N100m response generators during baseline. (b) Selective attention effect without the masker sound and during an easy task explained by a gain increase of feature-unspecific response and slightly sharper frequency tuning (which does not change the feature-specific response yet, as the probe sound is above threshold). (c) As in (b), but using continuous masker with wide notch. The feature-unspecific response is more strongly adapted when not paying attention than in (b), resulting in larger attentional modulation. (d) The case with narrow notch, where the feature-unspecific generator is strongly suppressed even during selective attention, but feature-specific generator still shows a prominent response. Attentional enhancement is in magnitude smaller, but proportionallly larger than in (c). (e) Lipreading effect explained by the shift of feature specificity of neurons. As the task is to discriminate /ba/ from /ga/, it is enough for the AC to discriminate the FM transitions. During lipreading /ga/, both generators are adapted more to the FM sweep matching /ga/. (f) Effect during lipreading vowels. The AC function is set to discriminate vowels, so the pure tones of all frequencies used as probes elicit weak responses.


a) Baseline, no top-down influence
Even sharper freq. tuning
Sharper freq.tuning
Frequency tuning
FM sweeps
Vowels
...
995
980
950
/a/
1000
1000
1000
/i/
1005
1020
1050
/o/
1010
1040
1100
/y/
"Best" feature
sharper frequency tuning
sharper tuning for higher-level features
0.5
Feature-specific generator
0.5
Feature-unspecific generator
1
Compound response
+
=
b) Attention to tones, no masker (SI and SII)
⇐ Small shift in preferred feature
AC response under baseline
AC response under various top-down influence
Attention effect ↑
c) Attention to tones, masker with wide notch and easy task (SI and SII)
Attention effect ↑
d) Attention to tones, masker with narrow notch and hard task (SI and SII)
Large shift in preferred feature
Strongly adapted by masker
Attention effect is proportionally larger than in c)
Attention effect ↑
e) Lipreading, easy task with CV syllables (SIII)
Task-specific shift in preferred feat.
Matching /ga/
Lipreading /ga/
Lipreading effect ↓
Lipreading /ga/ ↓↓
Strongly adapted by efference copy from the speech production system
f) Lipreading, easy task with vowels (SIV)
Large task-specific shift
Lipreading effect ↓

study showed a suppression effect only during normal auditory feedback compared to pitch-shifted feedback, suggesting specificity in the effect (Behroozmand et al., 2011). Pitch-shifted feedback actually increased the post-vocalization evoked responses to sound. Our results in PIV are in favor with the temporal dependency, although they show that the simultaneous timing of the signals only transiently modulates the ongoing suppression of task-irrelevant and incongruent sounds, with in magnitude smaller modulation than the main suppression effect. The anatomy of multisensory connections allow the fast and time-locked interactions, shown for instance by auditory-induced phase-locking of visual cortex oscillations in monkey (Lakatos et al., 2008) which could underlie the results of PIII and PIV as well. Confirming this would, however, require future studies designed to study the phenomena as the present results do not support it.

5.5 Potential of the noise maskers

Neuroscience studies have a long history of experiments where the subjects are presented a single salient picture or a sound and the brain activity during this isolated event is recorded. This can tell much of the basic mechanisms on how the stimulus is processed in the brain, but leads to a limited view of the brain functioning as in real-world situations the stimuli are rarely so well-controlled. When the stimulus is more ambiguous or closer to the threshold, the top-down influence such as selective attention can modify the perception more dramatically as a lot of effort is needed even for a simple stimulus detection. Our brain has to do more in order to process the minimal information from different senses, or information from several senses needs to be integrated like in the case of noisy environment when we benefit from seeing the speaker’s lips (Sumby and Pollack, 1954).

The deliberate reduction of signal-to-noise can be taken into extreme, adding noise so much that the stimulus is at the detection threshold, exactly what was done in PI and PII. Previously the notched-noise maskers have been widely used to estimate psychophysical tuning curves (Glasberg and Moore, 1990) and to study cochlear frequency processing as measured at the auditory brainstem (e.g., Abdala and Folsom, 1995). The notched-noise method to filter out or attenuate neural assemblies frequency selectively has since then, besides PI and PII, been used in several studies studying central auditory processing (Okamoto et al., 2007; Stracke et al., 2009; Okamoto et al., 2009, 2010; Ahveninen et al., 2011), thus showing the potential of the method in auditory attention studies. The notched-noise paradigm is comparable to informational masking setup (Gutschalk et al., 2008), where the attended target sounds are mixed together with randomly occuring masker sounds of different frequencies with a minimum spectral distance (analog to the notch width). These kind of masker paradigms can reveal fundamental mechanisms and frequency specificity of attention and other top-down influences in the field of auditory neuroscience.

5.6 Relation of the findings to speech perception

Taken to everyday life where sound and sight usually match, the results of PIII and PIV in part explain why we understand better what the person is speaking when we see the lip movements. Seeing the lips activates, at some level, our own speech production system, which in turn modifies our auditory cortex so that speech signal is analyzed more efficiently.

6. Conclusions

This Thesis mapped, at the gross level, neural mechanisms involved in the short-term auditory plasticity in humans. The use of EEG and MEG allowed studying the dynamics of the top-down modulations at millisecond level, while MEG combined with MRI images brought more precise information about the locus of the effects. The first two studies of auditory frequency selectivity (PI and PII) provided insight into basic auditory processing during selective attention. They showed that the attention increased the auditory cortex responses, with a combination of neural gain and selectivity increase, at two distinct stages following each other in time (PII). The lipreading-related studies (PIII and PIV) showed how visual interactions dynamically shape the auditory cortex function. A lipreading task not only suppressed the cortical responses, but did it in a feature specific way (PIII) depending on the lipread phoneme. This suppression is related to the speech production system, as shown by similarity of the effects during own speech production and lipreading (PIV).

The experimental data from this project, accompanied with models of top-down influence, helps us to explain neural mechanisms that underlie selective attention in everyday situations, for instance, how we start understanding speech in a noisy environment only after some time, and there, how seeing an articulating face can improve our understanding of heard speech.

Bibliography

Abbott, L. F. and Regehr, W. G. (2004). Synaptic computation. *Nature*, 431(7010):796–803.

Abdala, C. and Folsom, R. C. (1995). The development of frequency resolution in humans as revealed by the auditory brain-stem response recorded with notched-noise masking. *Journal of the Acoustical Society of America*, 98(2):921–930.

Ahlfors, S., Han, J., Belliveau, J., and Hämäläinen, M. (2010a). Sensitivity of MEG and EEG to source orientation. *Brain Topography*, 23:227–232.

Ahlfors, S. P., Han, J., Lin, F.-H., Witzel, T., Belliveau, J. W., Hämäläinen, M. S., and Halgren, E. (2010b). Cancellation of EEG and MEG signals generated by extended and distributed sources. *Human Brain Mapping*, 31(1):140–149.

Ahveninen, J., Hämälainen, M., Jääskeläinen, I. P., Ahlfors, S. P., Huang, S., Lin, F.-H., Raij, T., Sams, M., Vasios, C. E., and Belliveau, J. W. (2011). Attention-driven auditory cortex short-term plasticity helps segregate relevant sounds from noise. *Proceedings of the National Academy of Sciences of the United States of America*, 108(10):4182–4187.

Ahveninen, J., Jääskeläinen, I. P., Raij, T., Bonmassar, G., Devore, S., Hämälainen, M., Levänen, S., Lin, F.-H., Sams, M., Shinn-Cunningham, B. G., Witzel, T., and Belliveau, J. W. (2006). Task-modulated "what" and "where" pathways in human auditory cortex. *Proceedings of the National Academy of Sciences of the United States of America*, 103(39):14608–14613.

Alcaini, M., Giard, M.-H., Echallier, J.-F., and Pernier, J. (1994). Selective auditory attention effects in tonotopically organized cortical areas: A topographic ERP study. *Human Brain Mapping*, 2(3):159–169.

Alho, K., Medvedev, S. V., Pakhomov, S. V., Roudas, M. S., Tervaniemi, M., Reinikainen, K., Zeffiro, T., and Näätänen, R. (1999). Selective tuning of the left and right auditory cortices during spatially directed attention. *Cognitive Brain Research*, 7(3):335–341.

Alho, K., Töttölä, K., Reinikainen, K., Sams, M., and Näätänen, R. (1987). Brain mechanism of selective listening reflected by event-related potentials. *Electroencephalography and Clinical Neurophysiology*, 68(6):458–470.

Alho, K., Vorobyev, V. A., Medvedev, S. V., Pakhomov, S. V., Roudas, M. S., Tervaniemi, M., van Zuijen, T., and Näätänen, R. (2003). Hemispheric lateralization of cerebral blood-flow changes during selective listening to dichotically presented continuous speech. *Cognitive Brain Research*, 17(2):201–211.

Arnal, L. H., Morillon, B., Kell, C. A., and Giraud, A.-L. (2009). Dual neural routing of visual facilitation in speech processing. *The Journal of Neuroscience*, 29(43):13445–13453.

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

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

Behroozmand, R., Liu, H., and Larson, C. R. (2011). Time-dependent neural processing of auditory feedback during voice pitch error detection. *Journal of Cognitive Neuroscience*, 23(5):1205–1217.

Belin, P., Zatorre, R. J., Lafaille, P., Ahad, P., and Pike, B. (2000). Voice-selective areas in human auditory cortex. *Nature*, 403(6767):309–312.

Berger, H. (1929). Über das Elektroenkephalogramm des Menschen. *Archiv. für Psychiatrie und Nervenkrankheiten*, 87:527–570.

Besle, J., Fort, A., Delpuech, C., and Giard, M.-H. (2004). Bimodal speech: early suppressive visual effects in human auditory cortex. *European Journal of Neuroscience*, 20(8):2225–2234.

Bidet-Caulet, A., Fischer, C., Besle, J., Aguera, P.-E., Giard, M.-H., and Bertrand, O. (2007). Effects of selective attention on the electrophysiological representation of concurrent sounds in the human auditory cortex. *The Journal of Neuroscience*, 27(35):9252–9261.

Boly, M., Garrido, M. I., Gosseries, O., Bruno, M.-A., Boveroux, P., Schnakers, C., Massimini, M., Litvak, V., Laureys, S., and Friston, K. (2011). Preserved feedforward but impaired top-down processes in the vegetative state. *Science*, 332(6031):858–862.

Borg, E. and Zakrisson, J.-E. (1975). The activity of the stapedius muscle in man during vocalization. *Acta Oto-Laryngologica*, 79:325–333.

Bunzeck, N., Wuestenberg, T., Lutz, K., Heinze, H.-J., and Jancke, L. (2005). Scanning silence: Mental imagery of complex sounds. *NeuroImage*, 26(4):1119–1127.

Burrone, J. and Murthy, V. N. (2003). Synaptic gain control and homeostasis. *Current Opinion in Neurobiology*, 13(5):560–567.

Bushara, K. O., Hanakawa, T., Immisch, I., Toma, K., Kansaku, K., and Hallett, M. (2003). Neural correlates of cross-modal binding. *Nature Neuroscience*, 6(2):190–195.

Butler, R. A. (1968). Effect of changes in stimulus frequency and intensity on habituation of the human vertex potential. *Journal of the Acoustical Society of America*, 44(4):945–950.

Calvert, G. A., Bullmore, E. T., Brammer, M. J., Campbell, R., Williams, S. C. R., McGuire, P. K., Woodruff, P. W. R., Iversen, S. D., and David, A. S. (1997). Activation of auditory cortex during silent lipreading. *Science*, 276(5312):593–596.

Cappe, C. and Barone, P. (2005). Heteromodal connections supporting multisensory integration at low levels of cortical processing in the monkey. *European Journal of Neuroscience*, 22(11):2886–2902.

Cappe, C., Morel, A., Barone, P., and Rouiller, E. M. (2009). The thalamocortical projection systems in primate: An anatomical support for multisensory and sensorimotor interplay. *Cerebral Cortex*, 19(9):2025–2037.

Cariani, P. A. and Delgutte, B. (1996a). Neural correlates of the pitch of complex tones. I. Pitch and pitch salience. *Journal of Neurophysiology*, 76(3):1698–1716.

Cariani, P. A. and Delgutte, B. (1996b). Neural correlates of the pitch of complex tones. II. Pitch shift, pitch ambiguity, phase invariance, pitch circularity, rate pitch, and the dominance region for pitch. *Journal of Neurophysiology*, 76(3):1717–1734.

Cedolin, L. and Delgutte, B. (2005). Pitch of complex tones: Rate-place and interspike interval representations in the auditory nerve. *Journal of Neurophysiology*, 94(1):347–362.

Cherry, E. C. (1953). Some experiments on the recognition of speech, with one and with two ears. *Journal of the Acoustical Society of America*, 25(5):975–979.

Cohen, D. (1968). Magnetoencephalography: Evidence of magnetic fields produced by alpha-rhythm currents. *Science*, 161(3843):784–786.

Corbetta, M. and Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nature Reviews Neuroscience*, 3(3):201–215.

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

Dale, A. M., Fischl, B., and Sereno, M. I. (1999). Cortical surface-based analysis: I. Segmentation and surface reconstruction. *NeuroImage*, 9(2):179–194.

Dalley, J. W., McGaughy, J., O'Connell, M. T., Cardinal, R. N., Levita, L., and Robbins, T. W. (2001). Distinct changes in cortical acetylcholine and noradrenaline efflux during contingent and noncontingent performance of a visual attentional task. *The Journal of Neuroscience*, 21(13):4908–4914.

Dallos, P. (1992). The active cochlea. *The Journal of Neuroscience*, 12(12):4575–4585.

Debener, S., Ullsperger, M., Siegel, M., and Engel, A. K. (2006). Single-trial EEG-fMRI reveals the dynamics of cognitive function. *Trends in Cognitive Sciences*, 10(12):558–563.

Degerman, A., Rinne, T., Salmi, J., Salonen, O., and Alho, K. (2006). Selective attention to sound location or pitch studied with fMRI. *Brain Research*, 1077(1):123–134.

Delorme, A. and Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, 134(1):9–21.

Diehl, R. L., Lotto, A. J., and Holt, L. L. (2004). Speech perception. *Annual Review of Psychology*, 55(1):149–179.

Donald, M. W. and Young, M. J. (1982). A time-course analysis of attentional tuning of the auditory evoked response. *Experimental Brain Research*, 46(3):357–367.

Escera, C., Yago, E., Corral, M.-J., Corbera, S., and Nuñez, M. I. (2003). Attention capture by auditory significant stimuli: semantic analysis follows attention switching. *European Journal of Neuroscience*, 18(8):2408–2412.

Falchier, A., Clavagnier, S., Barone, P., and Kennedy, H. (2002). Anatomical evidence of multimodal integration in primate striate cortex. *The Journal of Neuroscience*, 22(13):5749–5759.

Fischl, B., Sereno, M. I., and Dale, A. M. (1999). Cortical surface-based analysis: II: Inflation, flattening, and a surface-based coordinate system. *NeuroImage*, 9(2):195–207.

Formisano, E., Kim, D.-S., Di Salle, F., van de Moortele, P.-F., Ugurbil, K., and Goebel, R. (2003). Mirror-symmetric tonotopic maps in human primary auditory cortex. *Neuron*, 40(4):859–869.

Fowler, C. A. and Saltzman, E. (1993). Coordination and coarticulation in speech production. *Language and Speech*, 36(2–3):171–195.

Fritz, J., Shamma, S., Elhilali, M., and Klein, D. (2003). Rapid task-related plasticity of spectrotemporal receptive fields in primary auditory cortex. *Nature Neuroscience*, 6(11):1216–1223.

Fritz, J. B., Elhilali, M., and Shamma, S. A. (2005). Differential dynamic plasticity of A1 receptive fields during multiple spectral tasks. *The Journal of Neuroscience*, 25(33):7623–7635.

Froemke, R. C., Merzenich, M. M., and Schreiner, C. E. (2007). A synaptic memory trace for cortical receptive field plasticity. *Nature*, 450:425–429.

Fujiwara, N., Nagamine, T., Imai, M., Tanaka, T., and Shibasaki, H. (1998). Role of the primary auditory cortex in auditory selective attention studied by whole-head neuromagnetometer. *Cognitive Brain Research*, 7(2):99–109.

Ghazanfar, A. A. and Schroeder, C. E. (2006). Is neocortex essentially multisensory? *Trends in Cognitive Sciences*, 10(6):278–285.

Giard, M.-H., Collet, L., Bouchet, P., and Pernier, J. (1994). Auditory selective attention in the human cochlea. *Brain Research*, 633(1–2):353–356.

Giard, M. H. and Peronnet, F. (1999). Auditory-visual integration during multimodal object recognition in humans: A behavioral and electrophysiological study. *Journal of Cognitive Neuroscience*, 11(5):473–490.

Glasberg, B. R. and Moore, B. C. J. (1990). Derivation of auditory filter shapes from notched-noise data. *Hearing Research*, 47(1–2):103–138.

Goldenholz, D. M., Ahlfors, S. P., Hämäläinen, M. S., Sharon, D., Ishitobi, M., Vaina, L. M., and Stufflebeam, S. M. (2009). Mapping the signal-to-noise-ratios of cortical sources in magnetoencephalography and electroencephalography. *Human Brain Mapping*, 30(4):1077–1086.

Grady, C., Van Meter, J. W., Maisog, J. M., Pietrini, P., Krasuski, J., and Rauschecker, J. P. (1997). Attention-related modulation of activity in primary and secondary auditory cortex. *NeuroReport*, 8(11):2511–2516.

Grill-Spector, K., Henson, R., and Martin, A. (2006). Repetition and the brain: neural models of stimulus-specific effects. *Trends in Cognitive Sciences*, 10:14–23.

Gutschalk, A., Micheyl, C., and Oxenham, A. J. (2008). Neural correlates of auditory perceptual awareness under informational masking. *PLoS Biology*, 6:e138.

Halpern, A. R. and Zatorre, R. J. (1999). When that tune runs through your head: A PET investigation of auditory imagery for familiar melodies. *Cerebral Cortex*, 9(7):697–704.

Hämäläinen, M. and Hari, R. (2002). Magnetoencephalographic characterization of dynamic brain activation: Basic principles and methods of data collection and source analysis. In Toga, A. W. and Mazziotta, J. C., editors, *Brain Mapping: The Methods (Second Edition)*, pages 227–253. Academic Press, San Diego.

Hämäläinen, M., Hari, R., Ilmoniemi, R. J., Knuutila, J., and Lounasmaa, O. V. (1993). Magnetoencephalography – theory, instrumentation, and applications to noninvasive studies of the working human brain. *Reviews of Modern Physics*, 65(2):413–497.

Hansen, J. C. and Hillyard, S. A. (1980). Endogeneous brain potentials associated with selective auditory attention. *Electroencephalography and Clinical Neurophysiology*, 49(3–4):277–290.

Hari, R. (1990). The neuromagnetic method in the study of the human auditory cortex. In Grandori, F., Hoke, M., and Romani, G. L., editors, *Advances in audiology*, volume 6. Auditory evoked magnetic fields and potentials, pages 222–282. Karger, Basel.

Hari, R., Hämäläinen, M., Kaukoranta, E., Mäkelä, J., Joutsiniemi, S.-L., and Tiihonen, J. (1989). Selective listening modifies activity of the human auditory cortex. *Experimental Brain Research*, 74:463–470.

Hari, R., Kaila, K., Katila, T., Tuomisto, T., and Varpula, T. (1982). Interstimulus interval dependence of the auditory vertex response and its magnetic counterpart: implications for their neural generation. *Electroencephalography and Clinical Neurophysiology*, 54(5):561–569.

Hari, R., Levänen, S., and Raij, T. (2000). Timing of human cortical functions during cognition: role of MEG. *Trends in Cognitive Sciences*, 4(12):455–462.

Hari, R. and Mäkelä, J. P. (1988). Modification of neuromagnetic responses of the human auditory cortex by masking sounds. *Experimental Brain Research*, 71:87–92.

Herrmann, K., Montaser-Kouhsari, L., Carrasco, M., and Heeger, D. J. (2010). When size matters: attention affects performance by contrast or response gain. *Nature Neuroscience*, 13(12):1554–1559.

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

Hillebrand, A. and Barnes, G. R. (2002). A quantitative assessment of the sensitivity of whole-head MEG to activity in the adult human cortex. *NeuroImage*, 16(3):638–650.

Hillyard, S. A., Hink, R. F., Schwent, V. L., and Picton, T. W. (1973). Electrical signs of selective attention in the human brain. *Science*, 182(4108):177–180.

Hillyard, S. A., Vogel, E. K., and Luck, S. J. (1998). Sensory gain control (amplification) as a mechanism of selective attention: Electrophysiological and neuroimaging evidence. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 393(1373):1257–1270.

Hirata, A., Aguilar, J., and Castro-Alamancos, M. A. (2006). Noradrenergic activation amplifies bottom-up and top-down signal-to-noise ratios in sensory thalamus. *The Journal of Neuroscience*, 26(16):4426–4436.

Houde, J. F., Nagarajan, S. S., Sekihara, K., and Merzenich, M. M. (2002). Modulation of the auditory cortex during speech: An MEG study. *Journal of Cognitive Neuroscience*, 14(8):1125–1138.

Huang, L. and Dobkins, K. R. (2005). Attentional effects on contrast discrimination in humans: evidence for both contrast gain and response gain. *Vision Research*, 45(9):1201–1212.

Hugdahl, K., Bronnick, K., Kyllingsbrk, S., Law, I., Gade, A., and Paulson, O. B. (1999). Brain activation during dichotic presentations of consonant-vowel and musical instrument stimuli: a ^{15}O-PET study. *Neuropsychologia*, 37(4):431–440.

Humphries, C., Liebenthal, E., and Binder, J. R. (2010). Tonotopic organization of human auditory cortex. *NeuroImage*, 50(3):1202–1211.

Jääskeläinen, I. P. (2010). The role of speech production system in audiovisual speech perception. *The Open Neuroimaging Journal*, 4(1):30–36.

Jääskeläinen, I. P., Ahveninen, J., Belliveau, J. W., Raij, T., and Sams, M. (2007). Short-term plasticity in auditory cognition. *Trends in Neurosciences*, 30:653–661.

Jääskeläinen, I. P., Ahveninen, J., Bonmassar, G., Dale, A. M., Ilmoniemi, R. J., Levänen, S., Lin, F.-H., May, P., Melcher, J., Stufflebeam, S., Tiitinen, H., and Belliveau, J. W. (2004a). Human posterior auditory cortex gates novel sounds to consciousness. *Proceedings of the National Academy of Sciences of the United States of America*, 101(17):6809–6814.

Jääskeläinen, I. P., Ojanen, V., Ahveninen, J., Auranen, T., Levänen, S., Möttönen, R., Tarnanen, I., and Sams, M. (2004b). Adaptation of neuromagnetic N1 responses to phonetic stimuli by visual speech in humans. *NeuroReport*, 15(18):2741–2744.

Jäncke, L., Mirzazade, S., and Shah, N. J. (1999). Attention modulates activity in the primary and the secondary auditory cortex: a functional magnetic resonance imaging study in human subjects. *Neuroscience Letters*, 266(2):125–128.

Jäncke, L. and Shah, N. J. (2004). ‘Hearing’ syllables by ‘seeing’ visual stimuli. *European Journal of Neuroscience*, 19(9):2603–2608.

Jones, E. G. (1998). Viewpoint: the core and matrix of thalamic organization. *Neuroscience*, 85(2):331–345.

Kaas, J. H. and Hackett, T. A. (2000). Subdivisions of auditory cortex and processing streams in primates. *Proceedings of the National Academy of Sciences of the United States of America*, 97(22):11793–11799.

Kandel, E. R., Siegelbaum, S. A., and Schwartz, J. H. (2000). Synaptic transmission. In Kandel, E. R., Schwartz, J. H., and Jessel, T. M., editors, *Principles of Neural Science*, 4th Ed., pages 175–186. McGraw-Hill, New York, NY.

Kayser, C. and Logothetis, N. K. (2009). Directed interactions between auditory and superior temporal cortices and their role in sensory integration. *Frontiers in Integrative Neuroscience*, 3:7.

Kerlin, J. R., Shahin, A. J., and Miller, L. M. (2010). Attentional gain control of ongoing cortical speech representations in a “cocktail party”. *The Journal of Neuroscience*, 30(2):620–628.

Klucharev, V., Möttönen, R., and Sams, M. (2003). Electrophysiological indicators of phonetic and non-phonetic multisensory interactions during audiovisual speech perception. *Cognitive Brain Research*, 18(1):65–75.

Knight, R. T., Hillyard, S. A., Woods, D. L., and Neville, H. J. (1981). The effects of frontal cortex lesions on event-related potentials during auditory selective attention. *Electroencephalography and Clinical Neurophysiology*, 52(6):571–572.

Knight, R. T., Scabini, D., and Woods, D. L. (1989). Prefrontal cortex gating of auditory transmission in humans. *Brain Research*, 504:338–342.

Knight, R. T., Staines, W. R., Swick, D., and Chao, L. L. (1999). Prefrontal cortex regulates inhibition and excitation in distributed neural networks. *Acta Psychologica*, 101(2–3):159–178.

Kraemer, D. J. M., Macrae, C. N., Green, A. E., and Kelley, W. M. (2005). Musical imagery: Sound of silence activates auditory cortex. *Nature*, 434(7030):158–158.

Kujala, T., Alho, K., and Näätänen, R. (2000). Cross-modal reorganization of human cortical functions. *Trends in Neurosciences*, 23(3):115–120.

Lakatos, P., Karmos, G., Mehta, A. D., Ulbert, I., and Schroeder, C. E. (2008). Entrainment of neuronal oscillations as a mechanism of attentional selection. *Science*, 320(5872):110–113.

Langner, G., Sams, M., Heil, P., and Schulze, H. (1997). Frequency and periodicity are represented in orthogonal maps in the human auditory cortex: evidence from magnetoencephalography. *Journal of Comparative Physiology A: Sensory, Neural, and Behavioral Physiology*, 181(6):665–676.

Leaver, A. M. and Rauschecker, J. P. (2010). Cortical representation of natural complex sounds: Effects of acoustic features and auditory object category. *The Journal of Neuroscience*, 30(22):7604–7612.

Lee, J. and Maunsell, J. H. R. (2009). A normalization model of attentional modulation of single unit responses. *PLoS ONE*, 4(2):e4651.

Lee, J. and Maunsell, J. H. R. (2010). Attentional modulation of MT neurons with single or multiple stimuli in their receptive fields. *The Journal of Neuroscience*, 30(8):3058–3066.

Levitt, H. (1971). Transformed up-down methods in psychoacoustics. *Journal of the Acoustical Society of America*, 49(2B):467–477.

Lewis, J. W. and Van Essen, D. C. (2000). Corticocortical connections of visual, sensorimotor, and multimodal processing areas in the parietal lobe of the macaque monkey. *Journal of Comparative Neurology*, 428:112–137.

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., and Studdert-Kennedy, M. (1967). Perception of the speech code. *The Psychological Review*, 74(6):431–431.

Liberman, A. M. and Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition*, 21(1):1–36.

Liberman, A. M. and Whalen, D. H. (2000). On the relation of speech to language. *Trends in Cognitive Sciences*, 4(5):187–196.

Lütkenhöner, B., Krumbholz, K., and Seither-Preisler, A. (2003). Studies of tonotopy based on wave N100 of the auditory evoked field are problematic. *NeuroImage*, 19(3):935–949.

MacSweeney, M., Amaro, E., Calvert, G. A., Campbell, R., David, A. S., McGuire, P., Williams, S. C., Woll, B., and Brammer, M. J. (2000). Silent speechreading in the absence of scanner noise: an event-related fMRI study. *NeuroReport*, 11(8):1729–1733.

Maison, S., Micheyl, C., and Collet, L. (2001). Influence of focused auditory attention on cochlear activity in humans. *Psychophysiology*, 38(1):35–40.

Massaro, D. W., Cohen, M. M., and Smeele, P. M. T. (1996). Perception of asynchronous and conflicting visual and auditory speech. *Journal of the Acoustical Society of America*, 100(3):1777–1786.

May, P. J. C. and Tiitinen, H. (2010). Mismatch negativity (MMN), the deviance-elicited auditory deflection, explained. *Psychophysiology*, 47(1):66–122.

McGurk, H. and MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, 264(5588):746–748.

Münte, T. F., Altenmüller, E., and Jäncke, L. (2002). The musician's brain as a model of neuroplasticity. *Nature Reviews Neuroscience*, 3(6):473–478.

Murray, S. O. and Wojciulik, E. (2004). Attention increases neural selectivity in the human lateral occipital complex. *Nature Neuroscience*, 7(1):70–74.

Musacchia, G., Sams, M., Nicol, T., and Kraus, N. (2006). Seeing speech affects acoustic information processing in the human brainstem. *Experimental Brain Research*, 168(1–2):1–10.

Musacchia, G., Sams, M., Skoe, E., and Kraus, N. (2007). Musicians have enhanced subcortical auditory and audiovisual processing of speech and music. *Proceedings of the National Academy of Sciences of the United States of America*, 104(40):15894–15898.

Näätänen, R. (1982). Processing negativity: An evoked-potential reflection of selective attention. *Psychological Bulletin*, 3:605–640.

Näätänen, R. (1992). *Attention and Brain Function*. Lawrence Erlbaum, Hillsdale, NJ.

Näätänen, R. and 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.

Näätänen, R., Teder, W., Alho, K., and Lavikainen, J. (1992). Auditory attention and selective input modulation: A topographical ERP study. *NeuroReport*, 3(6):493–496.

Nahum, M., Renvall, H., and Ahissar, M. (2009). Dynamics of cortical responses to tone pairs in relation to task difficulty: A MEG study. *Human Brain Mapping*, 30(5):1592–1604.

Neelon, M. F., Williams, J., and Garell, P. C. (2006). The effects of attentional load on auditory ERPs recorded from human cortex. *Brain Research*, 1118(1):94–105.

Neelon, M. F., Williams, J. C., and Garell, P. C. (2011). Elastic attention: Enhanced, then sharpened response to auditory input as attentional load increases. *Frontiers in Human Neuroscience*, 5:41.

Nelken, I. (2008). Processing of complex sounds in the auditory system. *Current Opinion in Neurobiology*, 18(4):413–417.

Nieuwenhuys, R. (1994). The neocortex. *Anatomy and Embryology*, 190:307–337.

Nishitani, N., Schurmann, M., Amunts, K., and Hari, R. (2005). Broca's region: From action to language. *Physiology*, 20(1):60–69.

Noesselt, T., Rieger, J. W., Schoenfeld, M. A., Kanowski, M., Hinrichs, H., Heinze, H.-J., and Driver, J. (2007). Audiovisual temporal correspondence modulates human multisensory superior temporal sulcus plus primary sensory cortices. *The Journal of Neuroscience*, 27(42):11431–11441.

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

Numminen, J., Salmelin, R., and Hari, R. (1999). Subject's own speech reduces reactivity of the human auditory cortex. *Neuroscience Letters*, 265(2):119–122.

O'Connell, M., Falchier, A., McGinnis, T., Schroeder, C., and Lakatos, P. (2011). Dual mechanism of neuronal ensemble inhibition in primary auditory cortex. *Neuron*, 69(4):805–817.

Ohshiro, T., Angelaki, D. E., and DeAngelis, G. C. (2011). A normalization model of multisensory integration. *Nature Neuroscience*, 14(6):775–782.

Okamoto, H., Stracke, H., Lagemann, L., and Pantev, C. (2010). Bottom-up driven involuntary auditory evoked field change: Constant sound sequencing amplifies but does not sharpen neural activity. *Journal of Neurophysiology*, 103(1):244–249.

Okamoto, H., Stracke, H., Wolters, C. H., Schmael, F., and Pantev, C. (2007). Attention improves population-level frequency tuning in human auditory cortex. *The Journal of Neuroscience*, 27(39):10383–10390.

Okamoto, H., Stracke, H., Zwitserlood, P., Roberts, L. E., and Pantev, C. (2009). Frequency-specific modulation of population-level frequency tuning in human auditory cortex. *BMC Neuroscience*, 10(1):1.

Pantev, C., Bertrand, O., Eulitz, C., Verkindt, C., Hampson, S., Schuierer, G., and Elbert, T. (1995). Specific tonotopic organizations of different areas of the human auditory cortex revealed by simultaneous magnetic and electric recordings. *Electroencephalography and Clinical Neurophysiology*, 94(1):26–40.

Parasuraman, R. and Beatty, J. (1980). Brain events underlying detection and recognition of weak sensory signals. *Science*, 210(4465):80–83.

Paus, T., Zatorre, R. J., Hofle, N., Caramanos, Z., Gotman, J., Petrides, M., and Evans, A. C. (1997). Time-related changes in neural systems underlying attention and arousal during the performance of an auditory vigilance task. *Journal of Cognitive Neuroscience*, 9(3):392–408.

Pekkola, J., Ojanen, V., Autti, T., Jääskeläinen, I. P., Möttönen, R., Tarkiainen, A., and Sams, M. (2005). Primary auditory cortex activation by visual speech: an fMRI study at 3 T. *NeuroReport*, 16(2):125–128.

Petkov, C. I., Kang, X., Alho, K., Bertrand, O., Yund, E. W., and Woods, D. L. (2004). Attentional modulation of human auditory cortex. *Nature Neuroscience*, 7(6):658–663.

Petkov, C. I., O'Connor, K. N., and Sutter, M. L. (2007). Encoding of illusory continuity in primary auditory cortex. *Neuron*, 54:153–165.

Poremba, A., Saunders, R. C., Crane, A. M., Cook, M., Sokoloff, L., and Mishkin, M. (2003). Functional mapping of the primate auditory system. *Science*, 299(5606):568–572.

Pugh, K. R., Shaywitz, B. A., Shaywitz, S. E., Fulbright, R. K., Byrd, D., Skudlarski, P., Shankweiler, D. P., Katz, L., Constable, R. T., and Fletcher, J. (1996). Auditory selective attention: An fMRI investigation. *NeuroImage*, 4(3):159–173.

Raij, T., McEvoy, L., Mäkelä, J. P., and Hari, R. (1997). Human auditory cortex is activated by omissions of auditory stimuli. *Brain Research*, 745(1–2):134–143.

Rauschecker, J. P. (1999). Auditory cortical plasticity: a comparison with other sensory systems. *Trends in Neurosciences*, 22(2):74–80.

Rauschecker, J. P. (2011). An expanded role for the dorsal auditory pathway in sensorimotor control and integration. *Hearing Research*, 271(1–2):16–25.

Rauschecker, J. P. and Scott, S. K. (2009). Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. *Nature Neuroscience*, 12:718–724.

Rauschecker, J. P. and Tian, B. (2000). Mechanisms and streams for processing of ”what” and ”where” in auditory cortex. *Proceedings of the National Academy of Sciences of the United States of America*, 97(22):11800–11806.

Rauschecker, J. P., Tian, B., and Hauser, M. (1995). Processing of complex sounds in the macaque nonprimary auditory cortex. *Science*, 268(5207):111–114.

Recanzone, G. H., Guard, D. C., and Phan, M. L. (2000). Frequency and intensity response properties of single neurons in the auditory cortex of the behaving macaque monkey. *Journal of Neurophysiology*, 83(4):2315–2331.

Remez, R., Rubin, P., Pisoni, D., and Carrell, T. (1981). Speech perception without traditional speech cues. *Science*, 212(4497):947–949.

Reynolds, J. H. and Heeger, D. J. (2009). The normalization model of attention. *Neuron*, 61(2):168–185.

Rif, J., Hari, R., Hämäläinen, M. S., and Sams, M. (1991). Auditory attention affects two different areas in the human supratemporal cortex. *Electroencephalography and Clinical Neurophysiology*, 79(6):464–472.

Rinne, T., Balk, M. H., Koistinen, S., Autti, T., Alho, K., and Sams, M. (2008). Auditory selective attention modulates activation of human inferior colliculus. *Journal of Neurophysiology*, 100(6):3323–3327.

Rinne, T., Pekkola, J., Degerman, A., Autti, T., Jääskeläinen, I. P., Sams, M., and Alho, K. (2005). Modulation of auditory cortex activation by sound presentation rate and attention. *Human Brain Mapping*, 26(2):94–99.

Rizzolatti, G. and Arbib, M. A. (1998). Language within our grasp. *Trends in Neurosciences*, 21(5):188–194.

Roberts, T. P., Ferrari, P., Stufflebeam, S. M., and Poeppel, D. (2000). Latency of the auditory evoked neuromagnetic field components: stimulus dependence and insights toward perception. *Journal of Clinical Neurophysiology*, 17(2):114–119.

Rockland, K. S. and Ojima, H. (2003). Multisensory convergence in calcarine visual areas in macaque monkey. *International Journal of Psychophysiology*, 50(1–2):19–26.

Roffler, S. K. and Butler, R. A. (1968). Factors that influence the localization of sound in the vertical plane. *Journal of the Acoustical Society of America*, 43(6):1255–1259.

Romani, G. L., Williamson, S. J., and Kaufman, L. (1982). Tonotopic organization of the human auditory cortex. *Science*, 216(4552):1339–1340.

Salmi, J., Rinne, T., Degerman, A., Salonen, O., and Alho, K. (2007). Orienting and maintenance of spatial attention in audition and vision: multimodal and modality-specific brain activations. *Brain Structure and Function*, 212:181–194.

Salmi, J., Rinne, T., Koistinen, S., Salonen, O., and Alho, K. (2009). Brain networks of bottom-up triggered and top-down controlled shifting of auditory attention. *Brain Research*, 1286:155–164.

Sams, M., Aulanko, R., Hämäläinen, M., Hari, R., Lounasmaa, O. V., Lu, S.-T., and Simola, J. (1991). Seeing speech: visual information from lip movements modifies activity in the human auditory cortex. *Neuroscience Letters*, 127(1):141–145.

Sams, M., Möttönen, R., and Sihvonen, T. (2005). Seeing and hearing others and oneself talk. *Cognitive Brain Research*, 23(2–3):429–435.

Sams, M. and Salmelin, R. (1994). Evidence of sharp frequency tuning in the human auditory cortex. *Hearing Research*, 75(1–2):67–74.

Sarter, M., Givens, B., and Bruno, J. P. (2001). The cognitive neuroscience of sustained attention: where top-down meets bottom-up. *Brain Research Reviews*, 35(2):146–160.

Scheich, H., Brechmann, A., Brosch, M., Budinger, E., and Ohl, F. W. (2007). The cognitive auditory cortex: Task-specificity of stimulus representations. *Hearing Research*, 229:213–224.

Schroeder, C. E. and Foxe, J. J. (2002). The timing and laminar profile of converging inputs to multisensory areas of the macaque neocortex. *Cognitive Brain Research*, 14(1):187–198.

Schürmann, M., Raij, T., Fujiki, N., and Hari, R. (2002). Mind's ear in a musician: Where and when in the brain. *NeuroImage*, 16(2):434–440.

Shams, L., Kamitani, Y., and Shimojo, S. (2000). Illusions: What you see is what you hear. *Nature*, 408(6814):788–788.

Shomstein, S. and Yantis, S. (2004). Control of attention shifts between vision and audition in human cortex. *The Journal of Neuroscience*, 24(47):10702–10706.

Shomstein, S. and Yantis, S. (2006). Parietal cortex mediates voluntary control of spatial and nonspatial auditory attention. *The Journal of Neuroscience*, 26(2):435–439.

Skipper, J. I., van Wassenhove, V., Nusbaum, H. C., and Small, S. L. (2007). Hearing lips and seeing voices: How cortical areas supporting speech production mediate audiovisual speech perception. *Cerebral Cortex*, 17(10):2387–2399.

Stein, B. E. and Meredith, M. A. (1993). *The merging of the senses*. MIT Press, Cambridge, MA.

Stracke, H., Okamoto, H., and Pantev, C. (2009). Interhemispheric support during demanding auditory signal-in-noise processing. *Cerebral Cortex*, 19(6):1440–1447.

Stufflebeam, S. M., Poeppel, D., Rowley, H. A., and Roberts, T. P. (1998). Perithreshold encoding of stimulus frequency and intensity in the M100 latency. *NeuroReport*, 9(1):91–94.

Sumby, W. H. and Pollack, I. (1954). Visual contribution to speech intelligibility in noise. *Journal of the Acoustical Society of America*, 26(2):212–215.

Talavage, T. M., Sereno, M. I., Melcher, J. R., Ledden, P. J., Rosen, B. R., and Dale, A. M. (2004). Tonotopic organization in human auditory cortex revealed by progressions of frequency sensitivity. *Journal of Neurophysiology*, 91(3):1282–1296.

Talsma, D. and Woldorff, M. G. (2005). Selective attention and multisensory integration: Multiple phases of effects on the evoked brain activity. *Journal of Cognitive Neuroscience*, 17(7):1098–1114.

Tiitinen, H., Sinkkonen, J., Reinikainen, K., Alho, K., Lavikainen, J., and Näätänen, R. (1993). Selective attention enhances the auditory 40-Hz transient response in humans. *Nature*, 364(6432):59–60.

Tramo, M. J., Cariani, P. A., Koh, C. K., Makris, N., and Braida, L. D. (2005). Neurophysiology and neuroanatomy of pitch perception: Auditory cortex. *Annals of the New York Academy of Sciences*, 1060(1):148–174.

Tzourio, N., El Massioui, F., Crivello, F., Joliot, M., Renault, B., and Mazoyer, B. (1997). Functional anatomy of human auditory attention studied with PET. *NeuroImage*, 5(1):63–77.

Ulanovsky, N., Las, L., and Nelken, I. (2003). Processing of low-probability sounds by cortical neurons. *Nature Neuroscience*, 6(4):391–398.

van Wassenhove, V., Grant, K. W., and Poeppel, D. (2005). Visual speech speeds up the neural processing of auditory speech. *Proceedings of the National Academy of Sciences of the United States of America*, 102(4):1181–1186.

van Wassenhove, V., Grant, K. W., and Poeppel, D. (2007). Temporal window of integration in auditory-visual speech perception. *Neuropsychologia*, 45(3):598–607.

Watkins, K. E., Strafella, A. P., and Paus, T. (2003). Seeing and hearing speech excites the motor system involved in speech production. *Neuropsychologia*, 41(8):989–994.

Weinberger, N. M. (1998). Physiological memory in primary auditory cortex: Characteristics and mechanisms. *Neurobiology of Learning and Memory*, 70:226–251.

Weinberger, N. M. (2004). Specific long-term memory traces in primary auditory cortex. *Nature Reviews Neuroscience*, 5(4):279–290.

Weinberger, N. M. (2007). Auditory associative memory and representational plasticity in the primary auditory cortex. *Hearing Research*, 229:54–68.

Weinberger, N. M., Miasnikov, A. A., and Chen, J. C. (2006). The level of cholinergic nucleus basalis activation controls the specificity of auditory associative memory. *Neurobiology of Learning and Memory*, 86(3):270–285.

Wessinger, C. M., VanMeter, J., Tian, B., Van Lare, J., Pekar, J., and Rauschecker, J. P. (2001). Hierarchical organization of the human auditory cortex revealed by functional magnetic resonance imaging. *Journal of Cognitive Neuroscience*, 13(1):1–7.

Westbrook, G. L. (2000). Seizures and epilepsy. In Kandel, E. R., Schwartz, J. H., and Jessel, T. M., editors, *Principles of Neural Science*, 4th Ed., pages 910–935. McGraw-Hill, New York, NY.

Winer, J. A. and Lee, C. C. (2007). The distributed auditory cortex. *Hearing Research*, 229(1–2):3–13.

Winkowski, D. E. and Knudsen, E. I. (2008). Distinct mechanisms for top-down control of neural gain and sensitivity in the owl optic tectum. *Neuron*, 60(4):698–708.

Woldorff, M. G., Gallen, C. C., Hampson, S. A., Hillyard, S. A., Pantev, C., Sobel, D., and Bloom, F. E. (1993). Modulation of early sensory processing in human auditory cortex during auditory selective attention. *Proceedings of the National Academy of Sciences of the United States of America*, 90(18):8722–8726.

Woldorff, M. G. and Hillyard, S. A. (1991). Modulation of early auditory processing during selective listening to rapidly presented tones. *Electroencephalography and Clinical Neurophysiology*, 79(3):170–191.

Woods, D. L. and Alain, C. (2001). Conjoining three auditory features: An event-related brain potential study. *Journal of Cognitive Neuroscience*, 13(4):492–509.

Woods, D. L., Stecker, G. C., Rinne, T., Herron, T. J., Cate, A. D., Yund, E. W., Liao, I., and Kang, X. (2009). Functional maps of human auditory cortex: Effects of acoustic features and attention. *PLoS ONE*, 4(4):e5183.

Zatorre, R. J., Mondor, T. A., and Evans, A. C. (1999). Auditory attention to space and frequency activates similar cerebral systems. *NeuroImage*, 10(5):544–554.

Aalto-DD 58/2012

ISBN 978-952-60-4619-8
ISBN 978-952-60-4620-4 (pdf)
ISSN-L 1799-4934
ISSN 1799-4934
ISSN 1799-4942 (pdf)

Aalto University
School of Science
Department of Biomedical Engineering and Computational Science

BUSINESS +
ECONOMY

ART +
DESIGN +
ARCHITECTURE

SCIENCE +
TECHNOLOGY

CROSSOVER

DOCTORAL
DISSERTATIONS