Jussi Nikki

# Processing of connected speech in the human brain: MEG observations

**Faculty of Electronics, Communications and Automation**

Thesis submitted for examination for the degree of Master of Science in Technology.

Espoo 11.11.2010

**Thesis supervisor:**

Prof. Mikko Sams

**Thesis instructors:**

Ph.D. Hannu Tiitinen

Ph.D. Patrick May

Aalto University
School of Science
and Technology

AALTO UNIVERSITY
SCHOOL OF SCIENCE AND TECHNOLOGY

ABSTRACT OF THE
MASTER'S THESIS

Author: Jussi Nikki

Title: Processing of connected speech in the human brain: MEG observations

Date: 11.11.2010 Language: English Number of pages:8+59

Faculty of Electronics, Communications and Automation

Department of Biomedical Engineering and Computational Science

Professorship: Cognitive technology Code: S-114

Supervisor: Prof. Mikko Sams

Instructors: Ph.D. Hannu Tiitinen, Ph.D. Patrick May

Speech is a compound that consists of a sequence of syllables. Ultimately, it can be seen as a simple time series. In speech comprehension this physical representation is robustly linked to its meaning. Auditory processing in the human brain is organized into parallel streams, each stream containing hierarchical structures presumably specialized in a distinct part of this process. In the present study, activation in the cortex was studied during listening to speech of varying intelligibility, under varying levels of attention and past experience.

The auditory evoked field (P1m, N1m, P2m and sustained field SF), elicited during the presentation of complete Finnish sentences, was studied using magnetoencephalography. Controlled reduction of intelligibility was achieved by reducing the number of bits to represent the signal samples. Contribution of memory-related processes was studied using consecutive presentations of degraded and good-quality sentences.

As a result of stimulus degradation, the amplitudes of the P1m, N1m, P2m and SF increased, while the latencies of the N1m and P2m decreased. These effects were bihemispheric and localized to the vicinity of the auditory cortex. The degraded sentences were better understood after listening to the original ones. In the passive condition, this resulted in increased activation within the inferior parietal regions during the P1m, within the right-hemispheric central superior parietal region during the N1m, and within the central and anterior superior temporal regions (AST) and the posterior inferior temporal region during the P2m. Higher activation within the AST was further linked to stimulus intelligibility. Attention was associated with increase of SF amplitude and lower activation within the central and posterior superior parietal and the central inferior temporal region.

Keywords: attention, auditory cortex, auditory evoked response, comprehensibility, human, magnetoencephalography, memory, N1m, P1m, P2m, speech, sustained field

AALTO-YLIOPISTO
TEKNILLINEN KORKEAKOULU

DIPLOMITYÖN
TIIVISTELMÄ

Tekijä: Jussi Nikki

Työn nimi: Luonnollisen puheen prosessointi ihmisaivoissa: MEG havaintoja

Päivämäärä: 11.11.2010 Kieli: Englanti Sivumäärä:8+59

Elektroniikan, tietoliikenteen ja automaation tiedekunta

Lääketieteellisen tekniikan ja laskennallisen tieteen laitos

Professuuri: Kognitiivinen teknologia Koodi: S-114

Valvoja: Prof. Mikko Sams

Ohjaajat: Ph.D. Hannu Tiitinen, Ph.D. Patrick May

Puhe muodostuu peräkkäisten äänteiden erilaisista yhdistelmistä. Pohjimmiltaan sen voidaan ajatella muodostavan yksinkertaisen aikasarjan, joka kuunneltaessa yhdistetään vankasti sen merkitykseen. Tämän tapahtumasarjan on osoitettu olevan aivoissa jakautunut useisiin erikoistuneisiin peräkkäisiin ja rinnakkaisiin osioihin. Tässä tutkimuksessa tarkasteltiin kuinka puheen ymmärrettävyys, siihen kohdistettu tarkkaavaisuus sekä aiempi kokemus heikkolaatuisen puheen sisällöstä vaikuttavat aivokuoressa tapahtuvaan aktivaatioon.

Lauseiden kuuntelun aiheuttamaa kuuloherätevastetta (P1m, N1m, P2m ja SF) tutkittiin käyttäen avuksi magnetoenkefalografiaa. Lauseiden laatua heikennettiin alentamalla ääninäytteissä käytettävien kvantisointitasojen lukumäärää. Muistiperäisten toimintojen osallisuutta tarkasteltiin vertailemalla heikkolaatuisten lauseiden aiheuttamaa aktivaatiota kun aiempaa kokemusta niiden sisällöstä ei ollut sekä kun ne olivat sisältönsä puolesta tuttuja.

Laadun heikennys johti molempien aivopuoliskojen kuuloaivokuorelta mitattujen P1m, N1m ja P2m vasteiden amplitudien kasvuun ja N1m ja P2m vasteiden viiveiden laskuun. Aiemman kokemuksen myötä heikkolaatuiset lauseet ymmärrettiin paremmin kuin alun perin, mikä johti muutoksiin aktivaatiossa passiivisen kuuntelun aikana. Aktivaatio lisääntyi P1m vasteen aikana päänlaenlohkon alaosissa, N1m vasteen aikana oikean puoleisen ylemmän päänlaenlohkon keskiosassa ja P2m vasteen aikana alemman ohimolohkon takaosassa ja ylemmän ohimolohkon keski- ja etuosissa. Viimeiseksi mainitun alueen korkeampi aktivaatio näyttää liittyvän lisäksi lauseiden ymmärrettävyyteen. Tarkkaavaisuus johti kuuloaivokuorelta mitatun SF vasteen amplitudin kasvuun ja alhaisempaan aktivaatioon alemman ohimolohkon keskiosassa ja ylemmän päänlaenlohkon keski- ja takasosissa.

Avainsanat: ihminen, kuuloaivokuori, kuuloherätevaste, magnetoenkefalografia, muisti, N1m, P1m, P2m, puhe, sustained field, tarkkaavaisuus, ymmärrettävyys

# Preface

This thesis was done in the Department of Biomedical Engineering and Computational Science (BECS) of the Aalto University School of Science and Technology. The supervisor of this thesis was Prof. Mikko Sams and the instructors were Ph.D. Patrick May and Ph.D. Hannu Tiitinen.

I would like to thank both my supervisor and my instructors for the opportunity to conduct this thesis at BECS. Additionally, I would like to express my gratitude to both of my instructors Ph.D. Hannu Tiitinen and Ph.D. Patrick May for their positive attitude and for their pure effort throughout the project. The directions and constructive proposals for improvement truly contributed to the final form of this thesis. Furthermore, I would like to thank Tapio Palomäki for data collection and Prof. Paavo Alku for his work on the auditory stimuli employed in this thesis. I would also like to thank my co-workers Santeri Yrttiaho, Ismo Miettinen and Nelli Salminen for their advice on statistical analyses and references and Jaakko Kauramäki for his valuable advice on MNE analyses.

Most of all, I would like to thank my partner Petra for her support especially during latter three months spent with this thesis.

Espoo, 11.11.2010 Jussi Nikki

# Contents

# Symbols and abbreviations

## Symbols

| | |
|---|---|
| $B$ | Magnetic flux density |
| $f_c$ | Filter cutoff frequency |
| $G_p$ | Gradient amplitude |

## Abbreviations

| | |
|---|---|
| **AEF** | Auditory evoked field |
| **AEP** | Auditory evoked potential |
| **AER** | Auditory evoked response |
| **ANOVA** | Analysis of variance |
| **dSPM** | Dynamic statistical parameter map |
| **ECD** | Equivalent current dipole |
| **EEG** | Electroencephalography |
| **EOG** | Electrooculography |
| **ERF** | Event-related field |
| **ERP** | Event-related potential |
| **ERR** | Event-related response |
| **FIR** | Finite impulse response |
| **fMRI** | Functional magnetic resonance imaging |
| **HPI** | Head position indicator |
| **MEG** | Magnetoencephalography |
| **MNE** | Minimum-norm estimate |
| **MR** | Magnetic resonance |
| **N1/N1m** | Negative AEP/AEF with peak latency of around 100 ms |
| **P1/P1m** | Positive AEP/AEF with peak latency of around 50 ms |
| **P2/P2m** | Positive AEP/AEF with peak latency of around 180 ms |
| **PET** | Positron emission tomography |
| **ROI** | Region of interest |
| **SEM** | Standard error of the mean |
| **SF** | Sustained field |
| **sLORETA** | Standardized low resolution brain electromagnetic tomography |
| **SNR** | Signal-to-noise ratio |
| **SP** | Sustained potential |
| **SPM** | Statistical parameter map |
| **SSS** | Signal-space separation |
| **STG** | Superior temporal gyrus |
| **STR** | Superior temporal region |
| **STS** | Superior temporal sulcus |
| **tSSS** | Spatio-temporal signal-space separation |
| **USQ** | Uniform scalar quantization |

## Abbreviations of ROI analysis regions

| | |
|---|---|
| **AIP** | Anterior inferior parietal region |
| **AIT** | Anterior inferior temporal region |
| **ASP** | Anterior superior parietal region |
| **AST** | Anterior superior temporal region |
| **CIP** | Central inferior parietal region |
| **CIT** | Central inferior temporal region |
| **CSP** | Central superior parietal region |
| **CST** | Central superior temporal region |
| **PIP** | Posterior inferior parietal region |
| **PIT** | Posterior inferior temporal region |
| **PSP** | Posterior superior parietal region |
| **PST** | Posterior superior temporal region |

# 1 Introduction

Human auditory processing has been studied extensively, the first documented experiments dating back to as early as 1863 when Herman von Helmholtz made an attempt to connect the physical properties of sounds to the perceived esthetics of music, eventually revealing the compound nature of all musical sounds (Davis & Merzbach 1975). The basic research performed during nearly one and a half century in numerous fields of science has provided us with a grounding that makes it possible to reduce large entities such as neural activity into quantifiable electric and magnetic fluctuations recorded in distinct locations of the brain. Current magnetoencephalography (MEG) and electroencephalography (EEG) instrumentation is capable of yielding information with sublime speed and temporal accuracy, enabling us to study the fluctuations with respect to events occurring during the experiments.

However, some limitations still remain. Neural activity can, quite naturally, be monitored only during the experiments, putting the emphasis on the experimental design itself. As the studies are constructed after the requirements of the instrumentation, to some degree they always represent merely a compromise between the real and the achievable - how the environment is perceived in the real world and to what extent we can artificially replicate it in laboratory conditions. To simulate our daily encounters, different stimuli have been used producing varied results. While using simplified stimuli often results in more uniform data, the findings obtained with more naturalistic stimuli can reveal effects that previously remained unseen. In the study of auditory cognition, this implies transformation from using stimuli such as tones of short duration to representing more meaningful vowels, syllables, or preferably, connected speech. Not surprisingly, as pure tones used already by von Helmholtz are certainly different from compound-like speech, the results have sometimes been non-overlapping. However, it would seem rational, that, when applied appropriately, naturalistic settings produce results that best represent the reality.

Human auditory processing has been focused on studying auditory evoked responses (AERs) that due to being temporally relatively static with respect to the stimuli, are easily reproducible. In other words, the information from the vast amount of recorded data is extracted using comparably small amount of features. As the presentation of a long-duration stimulus elicits a response with a length roughly that of the stimulus, one could assume that the sound is being processed during the same time frame. Regardless of this, many auditory studies have been examining only the beginning of the response, leaving the latter part relatively unexplored. This is not by any means wrong, and continues to be a justifiable decision. However, as the methods have advanced, it should be noted, that the limitation is no longer a prerequisite for the analyses.

Auditory processing in humans is organized into multiple serial streams, containing structures interpreting sounds of varying complexity. In the case of speech, in addition to the physical representation of the sound such as the time series of

the signal, the stimuli also conveys a lexical meaning that is vice versa linked to the physical properties. A known trait of human auditory perception is its robustness in linking these two ends - our capability to decipher the lexical meaning even when the stimulus heard is obscure and distorted, such as in noisy conditions. This phenomenon seems to be intertwined with two cognitive processes: attention and memory. Understanding the distorted stimulus requires knowing what to expect, but interestingly seems to require no attending to the sound. These factors offer an attractive opportunity to study the serial streams without altering the stimulus in any way: e.g. if the lexical meaning is missing, how will the stream be affected?

The goal of the present study was to investigate the mechanisms underlying our capability to understand speech under difficult circumstances, under varying levels of *a priori* knowledge and attention, by taking advantage of the lexical meaning conveyed by words in complete sentences. Providing the listeners with a foresight of the prospective sentences, made it possible to observe the memory-related processes that contribute to cortical auditory processing. To study the degree of involuntarity, both attentive and passive conditions were used. In addition to providing a more naturalistic perspective to auditory processing, the sentences, due to their length, made it possible to observe the auditory evoked sustained response.

# 2 Measuring cortical auditory processing using magnetoencephalography

## 2.1 Cortical auditory processing

### 2.1.1 The human auditory cortex

The human auditory cortex is located in the superior temporal region (STR) of the left and the right hemisphere. It represents the most central stage in the processing of auditory information along the auditory pathway. As deficiencies in the STR can lead to severe speech-related problems, such as inability to associate words with their meanings (verbal auditory agnosia) and inability to associate sounds with the events that produce them (nonverbal auditory agnosia), the whole area is thought to be essential for understanding speech (Budinger & Heil 2006). Essentially, the auditory cortex contains neurons that are responsive to acoustic stimulation. However, instead of consisting of only one homogenous field, the auditory cortex can be seen as a complex structure (Brugge & Reale 1985).

The anatomical and functional organization of the human auditory cortex is largely based on studies in non-human primates (Scott & Johnsrude 2003). These show that auditory cortex can be divided into three main parts; the core area which includes the primary auditory cortex, the belt area which surrounds the core area and the parabelt area which is located adjacent to the lateral belt along the lateral surface of the superior temporal gyrus (STG). An interesting property of these areas is hierarchical processing - the signals are first processed in the core then travel to the belt and then to the parabelt (Kaas et al. 1999). In monkey, the signals progress first through three core areas, which drive eight surrounding belt areas, which in turn, unlike the core areas, drive the parabelt areas (Kaas & Hackett 2000). Additionally, the belt and parabelt areas project information beyond the auditory region, to temporal, parietal and frontal lobes. (for a review, see: Scott & Johnsrude 2003). In addition to this feed-forward organization, also backward connections have been suggested to exits (Pandya 1995, Kaas & Hackett 2000, Scheich et al. 2007). The core areas are tonotopically organized, meaning that sounds with significantly different frequencies are processed parallelly in different processing streams.

Evidence of a similar hierarchical and parallel organization has been found also in the human auditory cortex. Using functional magnetic resonance imaging (fMRI), Wessinger et. al (2001) found that simple sounds such as pure tones activated primarily the core area, whereas the belt areas required more complex sounds to become activated. Using MEG, Inui et al. (2006) showed that activation progresses in distinct stages, probably reflecting a core-belt-parabelt organization. Similarly to the findings on the three core areas of a monkey, the human auditory cortex has been found to be organized in a tonotopic manner. Representation of stimuli of different frequencies elicit activation systematically in different locations of the primary auditory cortex, the higher frequency sounds resulting in activation in deeper parts

and lower frequencies being represented at more superficial parts of the auditory cortices (e.g. Romani et al. 1982, Pantev et al. 1995).

The left hemispheric auditory cortex is larger than the right hemispheric one. The difference in size may be linked to the left hemispheric specialization of the processing of speech (Geschwind & Levitsky 1968, Penhune et al. 1996). Also, based on source locations of auditory responses, the left auditory cortex is significantly more superior and slightly more posterior than the right auditory cortex (Reite et al. 1988). Supporting the theory of left-hemispheric specialization to speech processing, two speech areas, Broca's and Wernicke's areas, are in most people both located in the left hemisphere (Goldstein 2007, p. 297). Broca's area, involved in the production of speech, is situated in the frontal lobe and Wernicke's area, specialized in the comprehension of speech, is located in the temporal lobe posterior to the auditory cortex. However, regardless of the theoretical division between speech production and comprehension in Broca's and Wernicke's areas, and the suggested hemispheric specialization for verbal and nonverbal sounds, it should be noted that the reality is not as unambiguous. Production and perception of speech are likely to be intertwined (Scott & Johnsrude 2003). Both hemispheres have been reported to participate in language processing and both language associated areas acts as a part of a larger complex network contributing to normal language processing (for a review, see Kutas et al. 2007).

### 2.1.2 Event-related responses

Event-related responses (ERRs) are signals embedded in EEG and MEG that can be recorded and associated with some specific event such as some physical or mental occurrence. The responses are often referred to as event-related potentials (ERPs) or event-related field (ERFs). ERPs contain information on voltage fluctuations that is usually recorded between an electrode placed on the scalp and a reference electrode. ERFs are the magnetic counterparts of ERPs and contain information on the changes of the measured magnetic field. As the recorded EEG and MEG signals consist of small ERRs enveloped by comparably large background noise, to make the ERRs distinguishable from noise, single-trial responses are averaged using the information on the times when the specific type of stimuli was presented during the experiment. The averaged samples are selected so that they bear consistent temporal relation to the event (time-locking). The ERRs are assumed to be constant over trial and independent of the background noise. As the non-time-locked background noise is assumed to vary randomly between samples, the averaging should result in reduction of noise (Fabiani et al. 2007), the decrease of noise being proportional to the square root of the number of averaged samples (Kaukoranta et al. 1986). Auditory evoked potentials (AEPs) and auditory evoked fields (AEFs) are subclasses of ERRs, associated with an auditory event - such as a sound presented to the subjects. Instead of representing a change in the non-specific, or ongoing brain activity, these responses are believed to be generated independently due to auditory information processing (Mäkinen et al. 2005).

Presenting an auditory stimulus activates a large number of neurons along the auditory pathway and in cortex. This produces magnetic fields and electronic potentials that are measurable outside the skull (Hari et al. 1980). The presentation of an auditory stimulus results in a number of transient deflections in the evoked response. These deflections can be divided into three groups by their latency: the early responses occurring during the first 8 ms after the presentation of the auditory stimulus, middle latency responses occurring between 8 and 50 ms and the late responses occurring between 50 and 300 ms (Picton et al. 1974). If the stimulus duration is long enough, the transient deflections are followed by a baseline shift lasting to the end of the stimulus (Picton et al. 1978a). The shift is called sustained field (SF) in magnetic recordings and sustained potential (SP) in electrical recordings. The late transient responses and the sustained response vary with changes in the psychological significance of the stimuli, making them particularly interesting (Scherg 1989).

### 2.1.3 The late transient deflections

The late AEP/AEF comprises four waves: P1, N1, P2 and N2 (Picton et al. 1974). The most prominent of these are the N1 response - peaking at around 100 ms after the stimulus onset - and the P2 - peaking at 180 ms after the presentation of the stimulus (Scherg 1989). The magnetic responses P1m, N1m and the P2m (sometimes termed P50m, P100m and P180m, N1m is sometimes referred to also as M100) correspond to their electric counterparts P1, N1 and P2 (also often referred to as P50, P100 and P180), respectively.

The transient P1m reflects activity in the belt areas of the auditory cortex (Liégeois-Chauvel et al. 1994, Yvert et al. 2005). Also the N1m response has been shown to be generated at and around the primary auditory cortex (Hari et al. 1980, Pantev et al. 1990). More precisely, the N1 response has been linked to activation in belt and parabelt areas (Inui et al. 2006), the activation extending from Heschl's gyrus onto planum temporale, superior temporal sulcus, middle temporal gyrus and posterior parts of superior temporal gyrus (Jääskeläinen et al. 2004).

Since the N1 response can be divided into several components (Näätänen & Picton 1987), the recorded wave representing combined activation of several processes beginning as early as 60-80 ms after stimuli onset (Woods 1995), the peak should not be considered as an unitary event. As the activation during the N1 comprises the whole auditory cortex, the N1 could represent a final stage of activation in the feedforward organization of the auditory cortex, the following responses such as the P2 developing as a result of feedback activity between the auditory areas (May & Tiitinen 2010). Because the N1 can be time-locked both to the beginning and to the end of a stimulus, the response has previously been suggested to reflect cortical activity related to abrupt changes in the auditory environment (Hari et al. 1987). However, as the response is linked to activation at belt and parabelt areas which are

known to be sensitive to spectral and temporal properties of sounds, the generators of N1 are likely to have a more profound role in the analysis of auditory information (May & Tiitinen 2010).

### 2.1.4 The sustained response

In the case of a continuous stimuli of sufficient duration, the transient responses are followed by a sustained response. The sustained response is a baseline shift that lasts throughout the presentation of the stimulus and has the same polarity as the N1 response. The latency for the beginning of the sustained response is not as accurately defined as the latency for the transient responses. By subtracting the waveforms of AEP's recorded using sounds of different lengths, Picton et al. (1978a) calculated that the sustained response starts building between the N1 and P2 responses, on average at 150 ms. There is limited information about the origin of the sustained response. It has been proposed to be generated either at and around the primary auditory cortex (Hari et al. 1980, Pantev et al. 2004) or in belt and parabelt areas (Gutschalk et al. 2002, Okamoto et al. 2010). Supporting the latter view, unlike the transient responses, the sustained response is not very sensitive to manipulation of the stimuli, implying that the transient and sustained responses could have physiologically distinct generators (Picton et al. 1978a, Pantev et al. 1994).

Although the sustained potential was discovered in the human brain more than fifty years ago by Köhler and Wegener (1955), the number of studies focusing on the sustained response, possibly due to technical challenges, has been remarkably small compared to the studies on the transient responses (Pantev et al. 1994). Whereas the more robust transient responses can be observed even in single-trial responses, study of the sustained response requires some consideration such as careful selection of filtering parameters (Mäkinen 2006). As shown by Mäkinen (2002, p. 20) and May and Tiitinen (2010, p. 82 - 84), the use of a high pass filter may result in responses that, regardless of their true nature, always appear to be transient (the start and the end of a sustained response appears as transient deflections while the middle part diminishes). Also, as the sustained response has been estimated to start building before the P2 response reaches its peak (Picton et al. 1978a), the resulting cross-contamination can make distinguishing between the two responses difficult. Thirdly, as the sustained response does not result in a single observable peak, localization of the source generators and quantification of the response amplitude cannot be performed at one definitive time point.

### 2.1.5 Effects of attention on the auditory evoked responses

Attention is a theoretical concept that is used to describe how the selection between relevant and irrelevant sensory information is made (for a review, see Lachter 2004). In the auditory domain, the N1 has been suggested to be the earliest response that can be reliably linked to attention, while the P1, in contrast, had been suggested to reflect mainly processes modulated by the physical properties of the stimulus (Coles & Rugg 1995, cited by Hermann & Knight 2001). However, attention has been

shown to modulate also the P1 response (e.g. Woldorff & Hillyard 1991, Woldorff et al. 1993). This section aims to briefly review the effects of attention on amplitude and latency of the auditory evoked P1, N1, P2 and sustained responses and to list brain regions associated with attention.

The P1 has been shown to be sensitive to manipulation of attention. Using short tones as stimuli and both EEG and MEG, Woldorff & Hillyard (1991) and Woldorff et al. (1993) showed the stimuli to elicit higher P1 responses in active condition - that is, when the subject was actively attending to the clicks - than in the passive condition - when the subject was not attending to the sounds. Similarly, both the N1 and P2 responses exhibit higher amplitudes in the active condition (Picton & Hillyard 1974). Recently, similar effects on the magnetically recorded N1m were reported by Okamoto et al. (2010). While attention affects AERs under silent listening conditions, attention can also be linked to our capability to detect sounds when masked e.g. with noise (selective attention). Under noisy conditions (tones enveloped in continuous band-stop filtered noise), attention has been shown to increase both the amplitude and the latency of the N1 response (Kauramäki et al. 2007).

The sustained response is similarly enhanced during active listening. Picton et al. (1978a) found the sustained response to have a higher mean amplitude in the active condition than in the passive condition. However, the sustained response could still be observed in the passive condition (Picton et al. 1978a). Okamoto et al. (2010) replicated the results using MEG.

Information on areas involved in auditory attention is limited. Positron emission tomography (PET) and fMRI studies have proposed, several cortical regions to be involved in auditory attention (for a review, see Giard et al. 2000). These include the auditory cortex, thalamus, several regions of the frontal cortex, bilateral precentral and left postcentral cortices and the motor areas. In the auditory cortex, attention has been shown to enhance activation both in primary and secondary areas (Jäncke et al. 1999). Spatial directing of attention (i.e. attending to a sound with a specific location) has been associated with activation in parietal and frontal areas at around 180 ms and 200 ms, respectively, after the stimulus onset (Tiitinen et al. 2006). Studying the effects of multimodal (visual and auditory) attention, Salmi et al. (2007) found attention to enhance activation in the superior and inferior parietal cortex and the posterior prefrontal cortex.

### 2.1.6 Effects of sound degradation

The effects of stimulus parameters on the N1 response have been widely studied (Näätänen, Picton 1987, May & Tiitinen 2010). In this thesis, the emphasis is on such parameters that affect the intelligibility of the stimuli - whether the subject can understand the lexical meaning of the stimulus. The intelligibility of a stimulus can be varied using at least four distinct methods: reconstructing the natural

stimuli using sine-wave components (Miettinen et al. 2010), introducing continuous stochastic noise of varying sound pressure levels to mask the original stimuli (Martin et al. 1997, Whiting et al. 1998, Shtyrov et al. 1999, Martin & Stapells 2005, Morita et al. 2006, Kauramäki et al. 2007, Billings et al. 2009, Okamoto et al. 2010), using transient stochastic noise to mask the stimuli only during the epochs (Alho 2010), or re-quantizing the original stimuli using a lower amount of quantization levels (Liikkanen et al. 2007, Miettinen et al. 2010).

These studies have showed that the N1 response is highly sensitive to stimulus degradation. However, the trend of the results seems to vary with the method used to lower the intelligibility. Using continuous filtered stochastic noise and different stimuli (syllables, vowels or tones) Martin et al. (1997), Whiting et al. (1998), Shtyrov et al. (1999), Martin & Stapells (2005), Morita et al. (2006) and Billings et al. (2009) found that the N1 response decreases in amplitude and increases in latency when noise level is increased. Shtyrov (1999) found the decrease in amplitude to occur only in the left hemisphere. Okamoto et al. (2010) reported similar amplitude effects but no increase in latency. In contrast, using uniform scalar quantization (USQ) to degrade vowel sounds, Liikkanen et al. (2007) found that as a result, the amplitudes of N1 were increased in the right hemisphere. Later, using the same method, Miettinen et al. (2010) observed these effects in both hemispheres and also for the amplitudes of the P1 and P2 responses. In neither of the studies employing USQ, was sound degradation observed to have any latency effects (Liikkanen et al. 2007, Miettinen et al. 2010). The differences between the studies might relate to the fact that excluding the studies employing USQ, the stimuli were degraded using masks whose spectra did not correlate (Miettinen et al. 2010). This alters the resulting stimulus spectra and might also lower the effect of masking.

Recently, Alho (2010) conducted a study using transient noise, the mask being present only during the stimuli. This time, the results were congruent with the ones obtained with USQ by Liikkanen et al. (2007) and Miettinen et al. (2010), the N1 responses increasing in amplitude and decreasing in latency when the stimuli were degraded. In USQ, similarly, the mask is present only during stimulus. Therefore, the differences might be related to the adaptive nature of human auditory processing and being caused simply by the fact that the noise used in previous studies was present continuously during the experiment (Alho 2010).

Sound intensity is as one of the parameters that affect the intelligibility of a stimulus. Stimulus intensity affects the N1 response so that as intensity is decreased, the N1 amplitude also decreases whereas the N1 latency increases (Näätänen & Picton 1987). Similarly, both the sustained response (Picton et al. 1978b) and the P2 (Davis & Zerlin 1966) increase in amplitude as a function of stimulus intensity. Interestingly, these effects are analogous to those in studies where stimuli were masked using continuous noise.

### 2.1.7 Isolated speech sounds vs. connected speech

Studies in the auditory domain have traditionally, due to their controllable acoustic features, used simple non-speech stimuli, such as sinusoidal sounds (Alku et al. 1999). To use simple stimuli is a reasoned choice when studying the spectral properties or temporal dynamics on auditory processing. Speech perception can be considered as a fundamental part of auditory processing. To more reliably study auditory processing, isolated speech sounds such as vowels and syllables have been used. Moreover, when listening to connected (or natural) speech, a stream of sound is linked to a meaningful utterance. As attending to speech seems effortless or even involuntary, in the light of parallel and hierarchical structures identified in the cortical auditory areas, one could assume that speech-specific regions exist (for a review, see Scott & Johnsrude 2003). This section aims to review the studies, identifying specialized areas involved in the process of speech perception.

Hemodynamic studies (fMRI and PET), comparing the activation elicited by speech (containing phonetic information) and non-speech sounds (e.g. Belin et al. 2000, Binder et al. 2000, Scott et al. 2000, Vouloumanos et al. 2001, Belin et al. 2002, Narain et al. 2003), have reported higher activation in the vicinity of the superior temporal sulcus (STS) to be associated with the presentation of speech sounds. Additionally, Vouloumanos et al. (2001) and Scott et al. (2000) found the superior temporal gyrus (STG) to be affected. Both hemispheres seem to participate in speech processing. Belin et al. (2000, 2002) observed the enhanced activation due to speech bilaterally. Scott et al. (2000), Vouloumanos et al. (2001) and Narain et al. (2003) were able to observe this effect only in the left hemisphere. Binder et al. (2000) found the effect to be slightly more prominent in the left hemisphere, implying a subtle hemispheric specialization. The importance of STS to speech processing is supported by studies focusing on aphasia, a condition in which patients can understand speech perfectly but not repeat it. Aphasia has consistently been associated with deficiencies in the left posterior STS (Anderson et al. 1999, Axer et al. 2001).

To study how the presence of meaningful content affects cortical activation, Scott et al. (2000) and Narain et al. (2003) presented subjects stimuli that contained both speech and reversed speech. The latter stimuli containing phonetic structures as ordinary speech, but no logical meaning, was in both studies found to elicit significantly lower activation in the anterior parts of the left-hemispheric STS. Thus, while the posterior parts of the STS / STG seem to be involved in processing of general phonetic information, the anterior parts of the STS seem to be especially important for interpreting linguistic content, implying the presence of a specialized pathway for speech comprehension. Interestingly, as this effect was observed during passive listening (Scott et al. 2000, Narain et al. 2003), in speech comprehension, no attending to the sounds seems to be required.

## 2.2 MEG data analysis

### 2.2.1 Magnetoencephalography in a nutshell

MEG is a noninvasive neuroimaging technique that employs an array of coil sensors positioned over the scalp. These sensors, cooled down to a very low temperature, have no electric resistance and are therefore extremely sensitive to changes in the ambient magnetic field. Since change in neuronal activity is linked to changes in the electrical currents within the brain, which in turn induces a magnetic field, the measured fluctuations in the magnetic field can be used as a direct measure of neural activity. The underlying principles have been comprehensively explained by e.g. Hämäläinen et al. (1993), Hari (1999), Hämäläinen & Hari (2002) and most recently Hansen et al. (2010). Therefore, due to the aim of this thesis, instead of further describing the MEG instrumentation and mathematical background, this section focuses on briefly reviewing the essential characteristics of which one should be aware of when utilizing MEG in order to study cortical auditory processing.

Timing is essential for studying the dynamics of cortical activity. MEG provides information about the dynamics of activities of populations of neurons of the cerebral cortex with a submillisecond time resolution. The main advantage of MEG is that it combines the capability of localizing brain areas with reasonable spatial accuracy with the possibility to record signals with excellent temporal accuracy. Although EEG is capable of providing a similar time resolution, source localization using EEG is limited due to the electrical properties of the medium. Unlike electric signals, magnetic signals are minimally distorted by the different intervening layers such as the skull and the scalp. Additionally, in MEG, there is no need for a reference site. PET and fMRI are methods which provide high-resolution spatial information on the extent of the activation but little information on its temporal dynamics. Whereas PET and fMRI utilize indirect measures such as blood oxygenation or metabolism and provide information averaged over a long time window, MEG directly reflects neural activation. Also, while fMRI and PET studies rely on subtractions between conditions, in MEG studies such operations, although possible, are not required (Hari & Hämäläinen 2002, Lopes da Silva 2010).

The aforementioned advantages make MEG a preferable choice when studying cortical auditory processing. However, some limitations exist. As demonstrated by von Helmholtz (1853, cited by Hämäläinen et al. 1993, p.430), an electromagnetic field that is measured outside a conductor can originate from an infinite number of source configurations inside the conductor. Due to this fundamental problem, often referred to as the inverse problem, the fields measured outside the skull cannot unambiguously be tracked to specific locations. This shortcoming can be sidestepped by using models to introduce assumptions, such as the number of sources, thus finding the most-suitable solution. This is often referred to as source localization. Using an inverse model, however, always involves uncertainty, resulting in the so-called reasonable spatial accuracy which has been estimated to be a few millimeters in favorable conditions (Hämäläinen et al. 1993, p. 418, Lopes da Silva 2010). The

study presented in this thesis employs two different inverse models which are introduced in chapters 2.2.4 and 2.2.5.

Another limitation of MEG is that the magnetic field observable outside the skull is influenced by the orientation of the neurons with respect to the skull. MEG can detect only the magnetic fields that have a component orthogonal to the skull, i.e. those that are generated by neuronal currents having a tangentially oriented component with respect to the skull (Lopes da Silva 2010). Because of this, sources at the top of gyri cause fields that appear weaker than those resulting from activation in the fissures. However, in addition to the disadvantages, the information on the possible directions of propagation can be used to support tackling the inverse problem. This knowledge can be exploited by introducing *orientation constraints* that specify the amount of deviation that a direction of a wave is allowed to have compared to the normal of the model surface, resulting in greater source localization accuracy.

In addition to orientation, the depth of a source affects significantly the recorded signal. MEG systems incorporate several kinds of magnetic sensors, which have different properties when it comes to detecting signals from distant sources. However, as the sensitivity of a sensor to distant neural sources is in proportion to the sensitivity to distant noise sources, none of the sensors are well suited to detecting deep sources. As an example, a signal generated four centimeters inside the head recorded using the most sensitive of the sensors, the magnetometer, is ten times weaker than a similar signal generated on the cortical surface whereas using the least sensitive, the planar gradiometer, the signal has a strength of only 2 percent of the superficial signal (Hari 1999, p. 1109). Therefore, MEG is best suited for studying activity in the cerebral cortex.

Finally, MEG signals are believed to be primarily generated by synchronous activity of tens of thousands of neurons (Lopes da Silva 2010). Unlike hemodynamic methods (fMRI and PET), MEG is able to detect small changes in the synchrony within neuronal population. This however, also means that sustained neuronal activity, visible in fMRI and PET recordings, might go undetected in MEG (Salmelin & Kujala 2006). It can be estimated that in order to achieve a typically measurable magnetic field of 10 nAm (Hämäläinen et al. 1993, p. 424), perfectly synchronous activity of around fifty thousand neurons corresponding to an area of about 0.63 $mm^2$ is required (Lopes da Silva 2010). However, due to asynchrony and misalignment between the signals generated by the activations, up to 93 percent of the signal might be cancelled (Halgren et al. 2000 cited by Hämäläinen & Hari 2002), meaning that under more realistic assumptions the activated area is significantly larger. Hari et al. (1997) estimated that as little as 1% of synchronously activated cells can determine 97% of the total signal. Therefore, it is important to understand that the measurable MEG signal probably represent only a part of a much bigger picture.

### 2.2.2 Spatiotemporal signal-space separation

The biomagnetic signals measured in MEG are very weak compared to the surrounding environmental noise. Compared to the Earth's static magnetic field, the fields evoked by brain activity are estimated to be roughly one billion times weaker. A purpose-built magnetically shielded room can lower the environmental magnetic noise by a factor of one thousand or 60 dB at 10 Hz. The use of only gradiometers instead of the more noise-prone magnetometers can reduce noise by another 60 dB (Parkkonen 2010). However, these measures protect MEG only from disturbances generated by far-field sources and have no effect in suppressing disturbances originating inside the shielded room, from near-field sources such as the apparatus itself and even from the subject. Since even the beating of a subjects heart can produce signals up to 1000 times larger than the neuromagnetic signal of interest (Vrba 2002), it becomes obvious that more sophisticated methods are required.

Signal-space separation (SSS, Taulu et al. 2004) is a method for compensating these external interference and sensor artifacts. The SSS separates the measured signals into two distinct parts by their source locations: to the brain signals originating inside the MEG sensor array and to unwanted external disturbances arising outside of the sensor array. The functionality of the method is based on the fact that modern MEG systems include a generous amount of channels, providing therefore oversampling of the magnetic fields. This means that the external disturbances originating further away from the sensors should be present in the signals of several channels. Thus, knowing also the geometry of the channels, one can find the interferences by searching for spatial patterns (Elekta Neuromag 2006). Conveniently, the SSS method makes minimal assumptions, requiring no prior knowledge about the interferences (Taulu & Kajola 2005) and should provide shielding by attenuating sources within more than one meter by a factor of 150 (Taulu et al. 2005). As the SSS is based directly on Maxwell's equations, the method is sometimes called Maxwell filtering.

However, SSS is capable of suppressing only signals arising from outside the MEG sensor array and cannot therefore remove interferences arising within the patient (Elekta Neuromag 2006). The temporal extension to signal-space separation (tSSS, Taulu & Simola 2006) employs statistical analysis of time information to find and remove artifacts from nearby sources without altering the brain responses. In the absence of nearby artifacts, the outcome of the tSSS algorithm is equivalent to SSS (Taulu & Simola 2006). In a study of auditory-evoked responses, the method was demonstrated to efficiently remove interferences and even to provide good a enough signal-to-noise ratio for single-trial analyses, at the same time leaving the auditory evoked responses intact (Taulu & Hari 2009).

### 2.2.3 Vector summing

Planar gradiometer sensors are arranged in orthogonal pairs and measure the spatial derivatives, or gradients, of the magnetic field component normal to the sensor

surface. At any one location, this gradient is described by two quantities: amplitude and direction. Thus, the gradient amplitude measured using a gradiometer pair can be calculated for a single sample as:

$$G_p = \sqrt{\frac{\partial B_z}{\partial x}^2 + \frac{\partial B_z}{\partial y}^2}, \tag{1}$$

where $\frac{\partial B_z}{\partial x}$ and $\frac{\partial B_z}{\partial y}$ denote the two orthogonal derivatives of the field components normal to the sensor surfaces (Elekta Neuromag 2008). Raij et al. (2000) noted that vector summing can simplify the analysis of evoked responses especially when the location of the source current remains relatively stationary but the orientation of the source current varies strongly, as often occurs in highly convoluted cortical areas such as the primary auditory cortex (Penhune et al. 1996). Vector summing is illustrated in Figure 1.

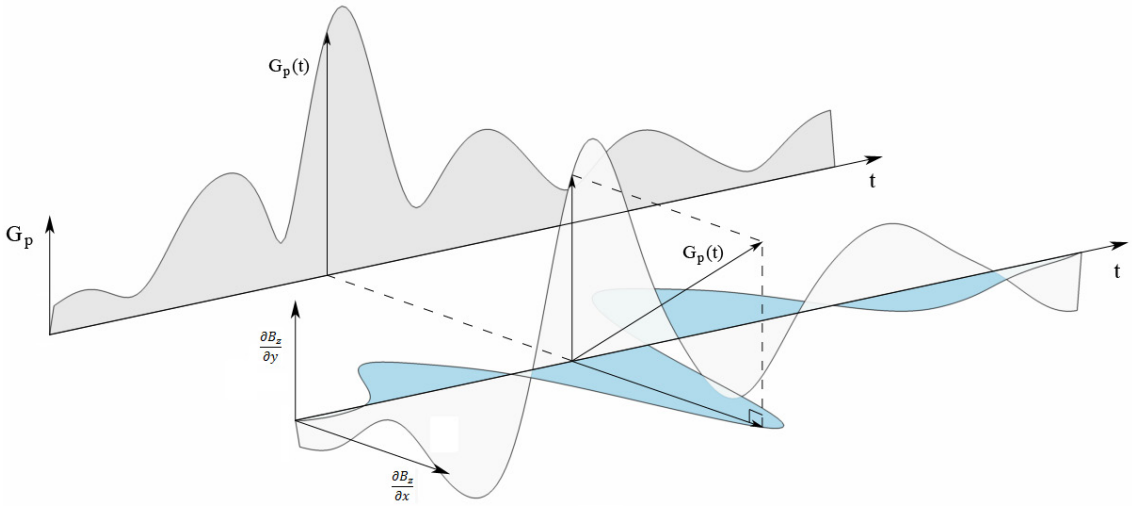


Figure 1: The principle underlying vector summing. The blue curve represents the $x$-component and light grey curve the $y$-component of the field derivative. The resulting amplitude $G_p$ is drawn in grey color.

### 2.2.4 Single equivalent current dipole modeling

Equivalent current dipole (ECD) modeling is an often used approach to locate the neural sources generating the measured magnetic field. In the simplest case, the magnetic field can be assumed to be generated by one point-like source. Since a localized cortical current source generates a magnetic-field which has a dipolar appearance (Brenner et al., 1978 cited by Hämäläinen et al. 1993), it is tempting to use this model to explain the generation of AEFs such as the N1m which usually have

dipolar field distributions. ECD-modeling is based on matching predicted signals and the measured signals in order to find the optimal solution and thus obtaining the dipole parameters - location, amplitude and direction of the source current. These parameters for the most suitable solution, commonly called the equivalent current dipole, or ECD, are found by minimizing the resulting least-square error between the measured and the predicted signals (Tuomisto et al. 1983).

To describe how well the modeled field of the ECD corresponds with the measured field and to validate the results, a goodness-of-fit value, $g$, can be used:

$$g = 1 - \frac{\sum_{i=1}^{n}(b_i - \hat{b_i})^2}{\sum_{i=1}^{n} b_i^2}, \tag{2}$$

where $n$ is the number of channels used for the modeling, $b_i$ are the measured field values and $\hat{b_i}$ are the field values modeled, i.e. those based on the ECD (Hämäläinen et al. 1993, p. 436). A value close to unity indicates a good fit. The use of a dipole can be considered to be justified when the value is over 0,65 (Kaukoranta et al. 1986). ECD modeling is usually performed using a subset of channels. In practice, non-biased location estimates require at least 20 channels. However, increasing the amount of channels will result in a lower $g$-value as more noise will be introduces along with the data.

Despite its simplicity, the use of the ECD model requires an understanding of its limitations. The simplest way of estimating the head conductivity for the computation of the magnetic field is to use a spherically symmetric conductor model. However, since radially directed current dipoles do not generate magnetic field outside this kind of conductor, only tangential dipoles are used (Kaukoranta et al. 1986). Also, since the ECD-model assumes that the signal is generated by a point-like source (or several sources in multi-dipole models), it is suited in studying activity only when it is possible to presume a fixed amount of sources and small source areas. If these assumptions are not valid, the results may be misleading (Hämäläinen & Ilmoniemi 1994). Because magnetic fields produced by multiple active regions and a single active region cannot easily be distinguished, the selection of the number of expected sources in multi-dipole models is difficult. Since also the requirement of a point-like source is often an oversimplification and the model does not provide any information on the extent of the activation, one should consider the ECD as an estimate on the center of gravity of the brain activity (Eggermont & Ponton 2002).

### 2.2.5 Minimum-norm current distribution estimates

Minimum-norm estimates (MNEs, Hämäläinen & Ilmoniemi 1994) are best estimates for source current distributions when a minimal amount of *a priori* informa-

tion about the source is available. Therefore, when calculating an estimate, the only assumption is to confine the source currents in a predefined region (Hämäläinen & Ilmoniemi 1994). The used region is the one producing the smallest norm of the primary current that is capable of explaining the measured signals (Hämäläinen et al. 1993). Thus, MNE is a method that finds the weakest possible current that could generate the magnetic field that was measured.

MNEs and ECDs represent two extreme approaches to source modeling. Whereas the MNE model uses minimal prior information on the source areas, the use of an ECD model can be justified only when it is know that only a small patch of the cortex produced the measured signals (Hämäläinen et al. 1993). The technique that the MNE employs, finding the most plausible source distribution, can provide estimates of complex source configurations without the complications of ECD modeling. However, even if the source were point-like, the estimate produced by MNE is blurred and has an extent of few centimeters in each dimension. To quantify the degree of blurring, a point-spread function can be used. It has been shown that the point-spread function of MNE has a location bias, i.e. inaccuracy of the estimates depends systematically on the location of the source (Dale et al. 2000). Consequently, the size of the activated region should not be used as an indicator of the true source area but should instead be regarded as an unavoidable limitation of the method (Hämäläinen & Hari 2002). In addition, MNEs fail to accurately measure the depth of the sources. Because the currents needed to explain the measured magnetic field are smaller if the sources are closer to the sensors, the MNE is biased towards solutions where the sources are more superficial (Hämäläinen & Hari 2002, Jensen & Hesse 2010).

### 2.2.6 Noise-normalized minimum-norm estimates

Noise-normalization is a method that aims to compensate for the inability of MNEs to accurately represent the true extent of activation. Noise-normalization is based on finding the areas where the recorded signal has an adequately high signal-to-noise ratio and subsequently delineating the noisy areas from the MNEs. The SNR judgement is made with the help of statistical parameter maps (or SPMs), also called dynamic statistical parameter maps (dSPMs) referring to a method by Dale et al. (2000). Importantly, the accuracy of the SPMs is not bound to the location of the source as strongly as it is with MNEs (Hämäläinen & Hari 2002), i.e. the point-spread function doesn't vary with the location of the source.

The dSPMs can be calculated by dividing the estimates of the current distribution resulting from the minimum-norm solutions by estimates of the noise standard deviation at each location and each time point (Dale et al. 2000). From these maps it is possible to find the areas where the estimated current value is significantly different from zero, indicating a good signal-to-noise ratio (Hämäläinen & Hari 2002). Currently, there are two noise-normalization methods that differ in the way the noise standard deviation estimates are calculated. Whereas the dSPMs (Dale et al.

2000) are calculated by using the standard deviation computed from the noise-only covariance, sLORETA (standardized low resolution brain electromagnetic tomography, Pascual-Marqui 2002) is a method employing the standard deviation calculated from the data covariance (Pantazis & Leahy 2010, p. 254-255). It has been shown that theoretically the sLORETA produces a zero location error (Pascual-Marqui 2002), making it preferable over dSPM. However, in realistic noise conditions the difference between the methods has been characterized as being less dramatic than claimed (Hämäläinen et al. 2010). The differences between estimates produced by the methods are illustrated in Figure 2.

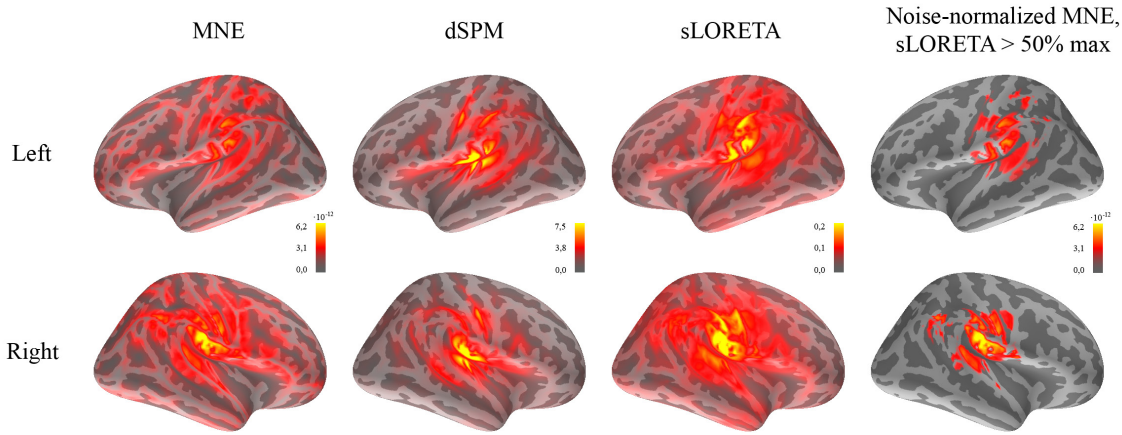


Figure 2: Estimates produced by the different methods visualized on the inflated surface of an average brain. Pictures show N1 response grand-averaged over the subjects. The data used for the simulation was obtained from the present study. The images were produced by averaging a 40-ms time window around the estimates of each subject's responses. The noise-normalized MNE was produced by displaying only those MNE voxels that have a value over 50 % of the maximum value in the sLORETA estimate. The scalings are set to show the maximum values of each type of estimate.

### 2.2.7 Interpreting the results of the minimum-norm estimates

Even though the point-spread function of the noise-normalized estimates (dSPM and sLORETA) is more uniform in space than that of MNE (Hämäläinen et al. 2010, p. 194), it should be noted that the roles of the MNE and the noise-normalized estimates are different. MNE provides an estimate of the current amplitude in physical units while the noise-normalized distributions are test statistics to be used as significance measures. Therefore, ideally the estimates should be used in combination, using the statistical significance map to delineate the regions of activity with high signal-to-noise ratio and consulting MNE for the true current amplitudes in those areas (Hämäläinen 2007, Hämäläinen et al. 2010).

To present the results as graphics, thresholding the displayed values is necessary. For the threshold value, which is used to crop the areas, it seems a common consen-

sus has not yet been reached. Employing sLORETA, Wagner et al. (2004) set the threshold at 50 % of the maximum value of the significance map. When locating regions of interests (ROIs), Dammers et al. (2007) used a threshold of 80 % of the maximal activity. Nummenmaa et al. (2007) remarked that the thresholding problem is relevant with all distributed inverse methods such as dSPM and sLORETA in which the prior variance is assumed to be uniform across the cortex such as the dSPM and the sLORETA. However, also they settled in setting the threshold "somewhat arbitrarily" to either display only a certain number of most relevant source locations or to include only sources which were above some percentage of the largest source amplitude (Nummenmaa et al. 2007). Halgren et al. (2002) applied dSPMs to create more focal representations from MNE estimates. However also this time, Jensen & Hesse (2010, p. 176) commented that the threshold applied in the study was somewhat arbitrary.

Gross et al. (2010) emphasize that it is important to understand the limitations of the methods especially when interpreting the results - most localization techniques cannot be reliably used to estimate the extent of an activated brain area. Even in the case of sophisticated statistical thresholding, the resulting area is often more related to some properties of the data such as signal-to-noise ratio than to the extent of the actual activation. (Gross et al. 2010). It should be noted that the quest for a perfect threshold value is actually not very valuable since the validation of the results should, as suggested by Hämäläinen (2007), be performed using the original MNE values.

## 2.3 Aims of the study

A known property of speech perception is its resilience to distortion. Humans are able to construct logical interpretations of signal which would be impossible to decipher with a computational algorithm. Instead of relying only on the acoustic features heard, one can assume that in humans this process is intertwined with memory - understanding a distorted stimulus requires knowing what to expect. If an initially incomprehensible degraded sentence was preceded by its intact correspondent, the degraded sentence should, anticipatedly, be better understood by the subject. Thus, if the subject first listened to a degraded sentence, then to its intact version, and finally to the very same degraded sentence again, the latter presentation of the degraded sentence would be physically identical to the first presentation while only the latter one would be comprehensible. This opens up an attractive possibility to study the underlying the memory-related processes and the effects of comprehensibility without altering the stimulus in any way.

The objectives of this study were to examine: (1) How does speech degradation itself affect the elicited cortical activation? (2) How does *a priori* information on the sentences affect activation and more specifically, to which region of the cortex are the memory-related processes localized? (3) In what way is attention involved in this process? The hypotheses for the current study were:

(a) The degraded sentences are initally difficult to comprehend.

(b) Listening to the intact sentences prior to the degraded sentences will contribute to understanding the degraded sentences.

(c) This contribution is introduced by memory-related processes and is reflected in the cortical activation.

Hypotheses (a) and (b) were futher tested with both subjective and objective measures.

# 3 Methods

## 3.1 MEG registrations

### 3.1.1 Auditory stimuli

The stimuli of the study were generated from Finnish sentences spoken by a native female speaker. The speaker produced altogether 84 sentences whose length varied from six to seven words and from 3 to 4.4 seconds. The structure of the sentences is presented in Figure 3. The sentences were constructed from three parts, each sentence representing a unique combination of the parts. The recordings of the data took place in the anechoic chamber of the Department of Signal Processing and Acoustics at Aalto University by utilizing a condenser microphone (Bruel&Kjaer 4188) and an AD-converter with a resolution of 16 bits and a sampling frequency of 22050 Hz. All the signals were high-pass filtered with a 6th order Butterworth filter (cut-off frequency at 60 Hz) to remove any low frequency fluctuation picked up by the microphone.

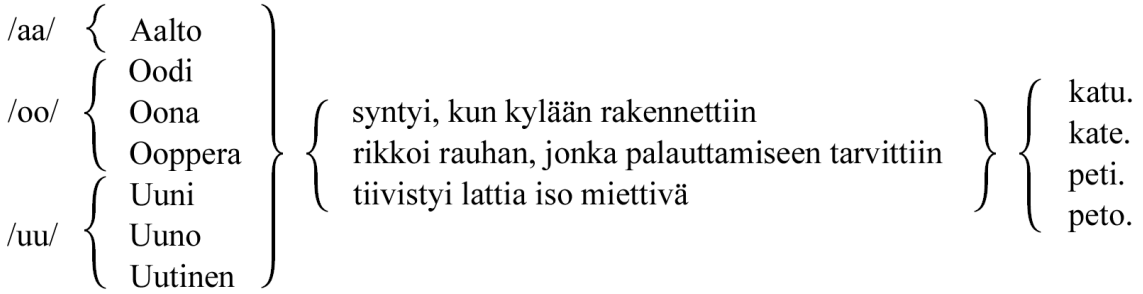


Figure 3: Structure of the sentences. Altogether 84 sentences beginning with either /aa/, /oo/ or /uu/ were generated.

In order to intentionally vary the perceptual quality of the spoken sentences, the data were processed with uniform scalar quantization (USQ, Cattermole, 1969). Controlled reduction of sound quality was achieved by decreasing the number of bits used in USQ from 16 to 1 bits. This resulted in two categories of sentences: natural sounding (16 bit) and degraded (1 bit). Before processing the signals with USQ, the maximum absolute signal value in the entire data set was identified. With this value, the signals were scaled so that the maximum absolute value corresponded to the largest positive level in the 16 bit scale (i.e. 32767). The amplitude scale adjusted in this manner was then used for all the sentences.

### 3.1.2 Experimental design

The experiment was designed to study the effects of three factors: sound degradation, attention and learning. These goals were achieved by taking advantage of the

lexical meaning of the sentences and the controlled reduction of sound quality provided by the USQ method. To study the effects of attention, both attentive active condition and passive ignore condition were used. To observe the effects of learning, consecutive presentations of reduced quality and good quality sentences were used.

The experimental design is illustrated in Figure 4. The experiment comprised of six sessions divided into active and passive parts (sessions 1-3 and 4-6 respectively). The active and passive parts were each further divided into three stimulus blocks, in which the first and third block consisted of the 1-bit-sentences, degraded using USQ and the intervening second block of good quality sentences (16-bit USQ). Each block comprised 120 sentences, selected from the 84 different ones, so that there was an equal amount of sentences starting with syllables /aa/, /oo/ and /uu/ (40 each).

During the first presentation of the degraded sentences (session 1), the subjects were expected to not understand the presented sentences, whereas due to the experience gained while listening the good-quality sounds, during the second presentation (session 3) the degraded sentences were expected to be better recognized. During the active listening condition (sessions 1 - 3) the ability of the subjects to understand the sentences was tested with both subjective and objective measures obtained from answers between sentences and questionnaires carried out between sessions. During the active condition, the trials consisted of a fixation cross, followed by an auditory stimulus (1 second interval), after which a question screen, inquiring whether the sentence was understood or not, was displayed. The passive condition comprised the same stimuli as the active condition. However, the participants were instructed to watch a silent movie of their choice while ignoring the auditory stimuli.

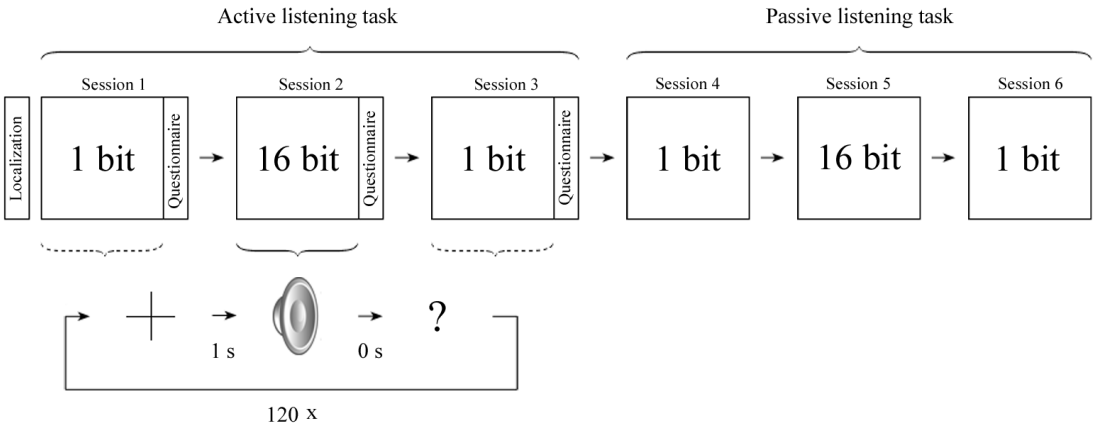


Figure 4: Experimental design. The experiment consisted of an active condition (sessions 1 - 3) and a passive condition (sessions 4 - 6). During the active condition, the ability of the subjects to comprehend the sentences was tested with questions.

### 3.1.3 Data acquisition

Ten right-handed volunteers (9 females, 1 male) participated in the study with informed consent. All of the participants were native Finnish speakers. The experiment was carried out in a magnetically shielded room at the Biomag Laboratory at Helsinki University Central Hospital. Auditory evoked fields were recorded using a 306-channel whole-head MEG system (Vectorview, Elekta Neuromag), the channels comprising 102 sensor elements, each containing one magnetometer and two orthogonally aligned gradiometers. The measured signals were sampled 600 times per second and low-pass filtered online with a cut-off frequency of 172 Hz. Additionally, for artifact rejection, the subject's eye movements were monitored using two pairs of EOG (Electrooculography) channels, one electrode pair set to measure horizontal and another vertical motion. In order to monitor head position, four head position indicator (HPI) coils were placed on the forehead and on the mastoids of the subject and localized within the MEG helmet using an Isotrak digitizer before the beginning of the first session. During the experiment, the subject was expected to remain sufficiently stationary. The auditory stimulus was presented binaurally through a pair of plastic tubes. The MEG signal was time-locked to the beginning of the auditory stimuli. The subjective questions, concerning the understanding of the stimuli, were presented visually and the answers inputted using two buttons placed under the index fingers, the left one implying a 'no' and the right one a 'yes' answer.

## 3.2 Data processing

### 3.2.1 Offline preprocessing

To identify recording artifacts and channels with noticeably lower signal-to-noise ratio (SNR), the recorded data was first visually inspected. Artifact-contaminated channels were omitted from later processing. After this, to suppress magnetic noise, spatio-temporal signal-space separation (Maxwell filtering with temporal extension) was performed for the data using Elekta Neuromag MaxFilter 2.0 software.

After spatiotemporal signal-space separation, two data sets with different filtering and averaging parameters were extracted from the data - one to study the transient responses, the P1, the N1 and the P2, and the other to study the complete AEF. The data for the analysis of the transient responses was filtered using a band-pass filter with a lower cutoff frequency of 2Hz (half-amplitude point) and an upper cutoff frequency of 30Hz. The finite impulse response (FIR) filter had a zero phase shift and transition band widths of 5Hz. For the analysis of the complete field, only low-pass filtering was performed. The filter had same characteristics as the low-pass filter used to process the previous data set.

After filtering, the data sets were corrected using a 100 ms pre-stimulus baseline. For averaging, the epochs were time locked to the beginning of the first word of the sentence using information recorded on the trigger channel. Epochs in which the peak-to-peak amplitude recorded using any of the gradiometer channels exceeded

2000 fT/cm were omitted from the average. For both of the EOG channel pairs, monitoring subjects' eye movements, the rejection threshold was 150 $\mu$V. For the analysis of the transient responses, to maximize the amount of artifact free epochs, the epochs were extracted using a 500 ms window. For the analysis of the complete auditory evoked field, a window of 3000 ms post-stimulus was used.

Due to an excessive amount of artifacts related to eye activity and a weak SNR, the data for one subjects was discarded from all further analyses. In addition, the data for one subject had to be omitted from the analysis of the SF due to an insufficient amount of averaged samples (i.e., 11 averaged epochs per session compared to an average of 81 epochs per session for the other subjects).

### 3.2.2 Latency and amplitude

The latencies and amplitudes of the transient responses were studied using amplitude values derived from the averaged gradiometer data. To maintain focus on the auditory areas of the brain, this was done by using a pair of gradiometers exhibiting the most prominent N1m response within 140 ms after stimulus onset. In order to perform the search, gradient amplitude distributions were calculated for the preprocessed and averaged data sets using equation 1 (p. 13). The analysis was carried out separately for each session, subject, and hemisphere and performed using MathWorks Matlab R2009b software.

The peaks of N1m, P1m and P2m responses were obtained using a semi-automatic approach. The peaks were located for each response by selecting a single time point with the highest local maximum within an expected time range. The P1m was expected to be found within 30 - 80 ms, the N1m within 90 - 140 ms and the P2m within 150 - 250 ms after stimulus onset. In cases where the peaks of responses proved to be difficult to locate, the gradiometer data was used to support the selection. Providing information on the direction of the field, the corresponding gradiometer channels proved to be helpful in selecting the time points in unclear cases. If a local maximum was not found within the expected latency in the calculated gradient amplitude, the data from the individual gradiometers was manually searched for prominent extreme values within the same time range.

### 3.2.3 Source localization

The magnetometer channels, being sensitive to homogenous fields often caused by distant noise sources (Hämäläinen & Hari 2002, p. 233), were discarded from the source analysis. Subsets of 44 gradiometer channels centered around the left and the right auditory cortices were used. The same selection has also previously been used to delimit the number of channels used for source modeling (see e.g. Huotilainen 1997, Miettinen et al. 2010, Liikkanen et al. 2007, Aho 2010). This kind of selection is large enough to produce unbiased estimates of the source locations in each auditory cortex, while containing no contribution from the opposite hemisphere.

To study the source locations of the found transient responses, single ECDs were fitted to the averaged data sets using Elekta Neuromag xfit Source Modeling Software v5.5. For modeling the conductivity of the head, a spherically symmetric model was used. Due to differences in the sizes and shapes of the subjects' heads, an accurately defined head position is essential for performing reliable dipole fitting. Before the first session, the head position of each subject was estimated with the help of points digitized from the head surface. Prior to ECD modeling, a sphere was fitted to these points in order to estimate the radius and the location of the origin of the sphere model. The coordinate system used in ECD modeling is defined in Figure 5. The $x$-axis passes through the two preauricular points and has a direction from left to right, $y$-axis passes through the nasion and points to the front. The $z$-axis points upwards.

A common approach is to fit dipoles sequentially, at short intervals within the expected time range of the response and to choose the dipole that results in the largest absolute dipole moment and an appropriate orientation of the dipole moment vector (e.g. Miettinen et al. 2010, Liikkanen et al. 2007, Alho 2010). However, in some cases, such as inadequate SNR, this kind of approach may prove unfeasible due to problems in identifying clear local maxima from the resulting absolute dipole moment values, inevitably leading to discarding a part of data from the analysis. A more robust solution is to use pre-existing information on the response latencies. In the current analyses, the dipoles were fitted at the peak latencies of the P1m, the N1m and the P2m responses (see 3.2.2), for each hemisphere separately.

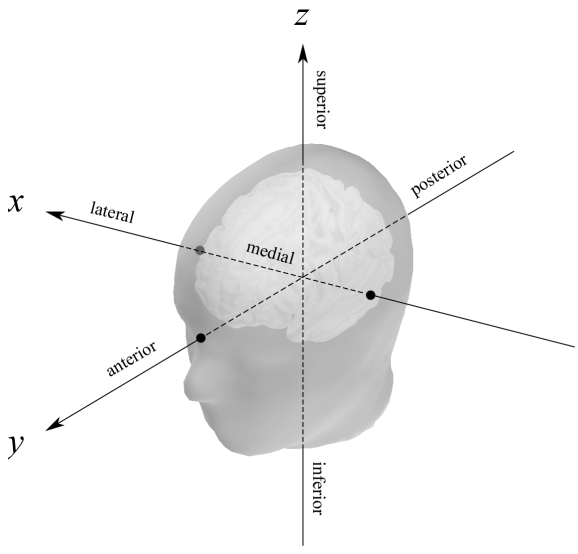


Figure 5: Coordinate system and anatomical terms of locations as used in ECD modeling. The $x$-axis passed through the two preauricular points and had a direction from left to right, $y$-axis passed through the nasion and pointed to the front. The $z$-axis pointed upwards.

### 3.2.4 The mean amplitude of the sustained field

After filtering and averaging, two time intervals could be distinguished in the SF of the AEF waveform (see Figure 16, page 36). The SF began as a large negative deflection that started at approximately 300 ms, reached its peak between 500 - 700 ms and ended at around 1000 ms after stimulus onset. For the rest of the stimulus duration, the amplitude of the SF remained relatively static.

As the nature of sustained response is not clear and it is thus difficult to identify the components of which the responses consist of, a recommended approach is to take a simple mean amplitude measurement over the waveform over a defined time interval (Picton et al. 2000). Thereby, mean amplitudes ($\overline{G_p}$) were calculated over the two intervals of 300 - 1000 ms and 1000 - 3000 ms using equation 3:

$$\overline{G_p} = \frac{1}{t_1 - t_0} \sum_{t=t_0}^{t_1} \sqrt{\frac{\partial B_z}{\partial x}(t)^2 + \frac{\partial B_z}{\partial y}(t)^2}, \tag{3}$$

where $\frac{\partial B_z}{\partial x}$ and $\frac{\partial B_z}{\partial y}$ denote the two orthogonal derivatives of the field components normal to the sensor surfaces and $t_0$ and $t_1$ the interval limits. The analysis was carried out separately for each hemisphere using the previously selected gradiometers displaying the highest N1m response amplitude. Calculations were made using the low-pass filtered data.

### 3.2.5 Current distribution estimates

To study activation spread in cortex, noise-normalized MNEs were calculated using the low-pass filtered data sets with the MNE Software (v2.7, Matti Hämäläinen). Noise-covariance matrices were calculated using the individual epochs of the Maxwell-filtered raw data files. Only the baseline period of 100 ms pre-stimulus was employed to the calculations. The forward solutions and the inverse operators were calculated for each session separately. For this purpose, a boundary-element model computed using average head and skull surface reconstructions (fsaverage) provided with the MNE software was used. Due to the average surface used, no orientation constraints were applied. Two different estimates, MNE and sLORETA, were calculated at intervals of 5 ms using the averaged datasets of each subject individually. To study the differences in current distribution during the transient responses in different conditions, both of the estimates were further averaged over 40-ms time windows, centered around the peaks of the responses. The latencies for these peaks, separately for each hemisphere of each subject, were obtained from previous analyses of transient responses (see 3.2.2). The two distinguishable periods of the SF were averaged in a similar fashion, however by using fixed time windows of 300 - 1000 ms for the early part and 1000 - 3000 ms for the late part of the SF. Finally, in order to perform noise-normalization, the calculated MNE estimates were delineated using

the values in the sLORETA estimates. The threshold value was set to 50 % of the maximum value of each estimate separately. The averaging was performed using Matlab software and the results were visualized using MNE software.

### 3.2.6 Region of interest analysis

To study the effects of attention and stimulus degradation on the cortical current distribution during the transient responses and the SF, the left- and the right-hemispheric cortices were each divided into 12 separate regions of interest (ROIs), six regions covering the temporal and six regions covering the parietal areas of the cortex. The regions were labeled according to their physical location: anterior, central or posterior and inferior or superior temporal/parietal region. The ROIs used for the analysis are illustrated in Figure 6. The mean current was calculated for each ROI using the MNE averages already calculated over the 40-ms time windows around the P1m, N1m and P2m peaks and over the two time periods of the SF. Since MNEs contain information on the current at each voxel of the cortical surface, the mean currents were obtained by calculating an average over all of the voxels belonging to each ROI. As suggested by Hämäläinen (2007), the MNE values were extracted from to original MNEs without noise-normalization. The labels for the ROIs were selected by hand using MNE software and the calculations were further done using Matlab software.

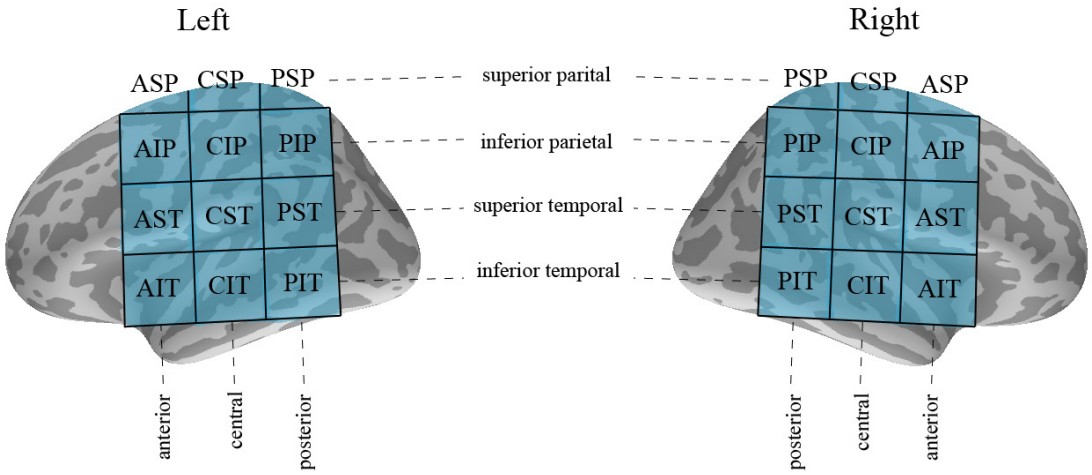


Figure 6: The regions of interest used for calculating the mean currents. The regions were labeled according to their physical locations. The first letter of the abbreviations refer to anterior/central/posterior, the second letter to inferior/superior and the third letter to parietal/temporal region

### 3.2.7 Statistical confirmation

The effects of the independent variables on the latencies, amplitudes, source location coordinates, mean amplitudes, and mean currents were studied using three-way

repeated-measures analysis of variance (ANOVA). The studied factors were hemisphere (left vs. right), attention (active vs. passive listening condition), and stimulus degradation (1 bit $1^{st}$ presentation vs. 16 bit vs. 1 bit $2^{nd}$ presentation). For all of the statistically significant results, Newman-Keuls post-hoc tests were performed. The calculations were done using Statistica software (v7.1, StatSoft).

# 4 Results

## 4.1 Identification accuracy

During the experiment, the subjects were first presented with sentences of degraded sound quality, followed by the same set of sentences in good sound quality. Finally, the subjects heard again the set of degraded sentences. According to hypothesis (b), during the second presentation, the participants would understand the degraded sentences better. This was tested during the active condition with questions in between the sentences. The identification percentage, that is, the proportion of sentences the subject reported having completely understood, rose significantly between the first and the second presentation of the degraded sentences for all of the subjects (see Figure 7). The mean identification percentages over the subjects were for the first and the second presentation 30.2 % (SEM: ± 7.6 %) and 78.9 % (± 3.7 %) respectively, indicating a remarkable increase of 48.7 percentage points ($F[1, 7] = 43.26$, $p < 0.0005$). In the case of the good quality sentences, the mean identification percentage was 94.8 % (± 0.9 %).

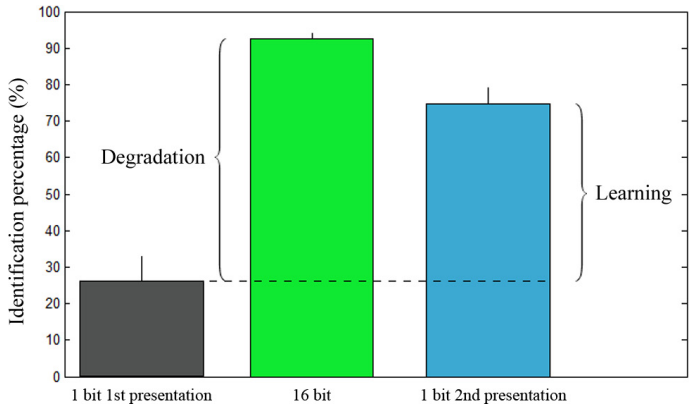


Figure 7: Mean identification percentage during the active listening condition. Error bars indicate standard error of the mean (SEM).

## 4.2 MEG observations

### 4.2.1 Amplitude and latency of the transient response

The P1m, N1m and P2m responses, elicited at the beginning of each sentence, were clearly observable in the waveforms of the auditory evoked fields, the P1m peaking on average at 66.3 ms, the N1m at 117.9 ms, and the P2m at 181.2 ms. The P1m had a mean amplitude of 21.9 fT/cm, the N1m 55.1 fT/cm and the P2m 31.3 fT/cm. Figure 8 shows the beginning part of the grand-averaged AEFs from the left- and right-hemispheric gradiometer channels displaying the maximum N1m response amplitudes. For the complete AEF, see Figure 16 on page 36. The mean amplitudes and the latencies of the P1m, N1m and P2m in the different conditions and stimulus types are shown in Figures 9, 10 and 11 respectively.

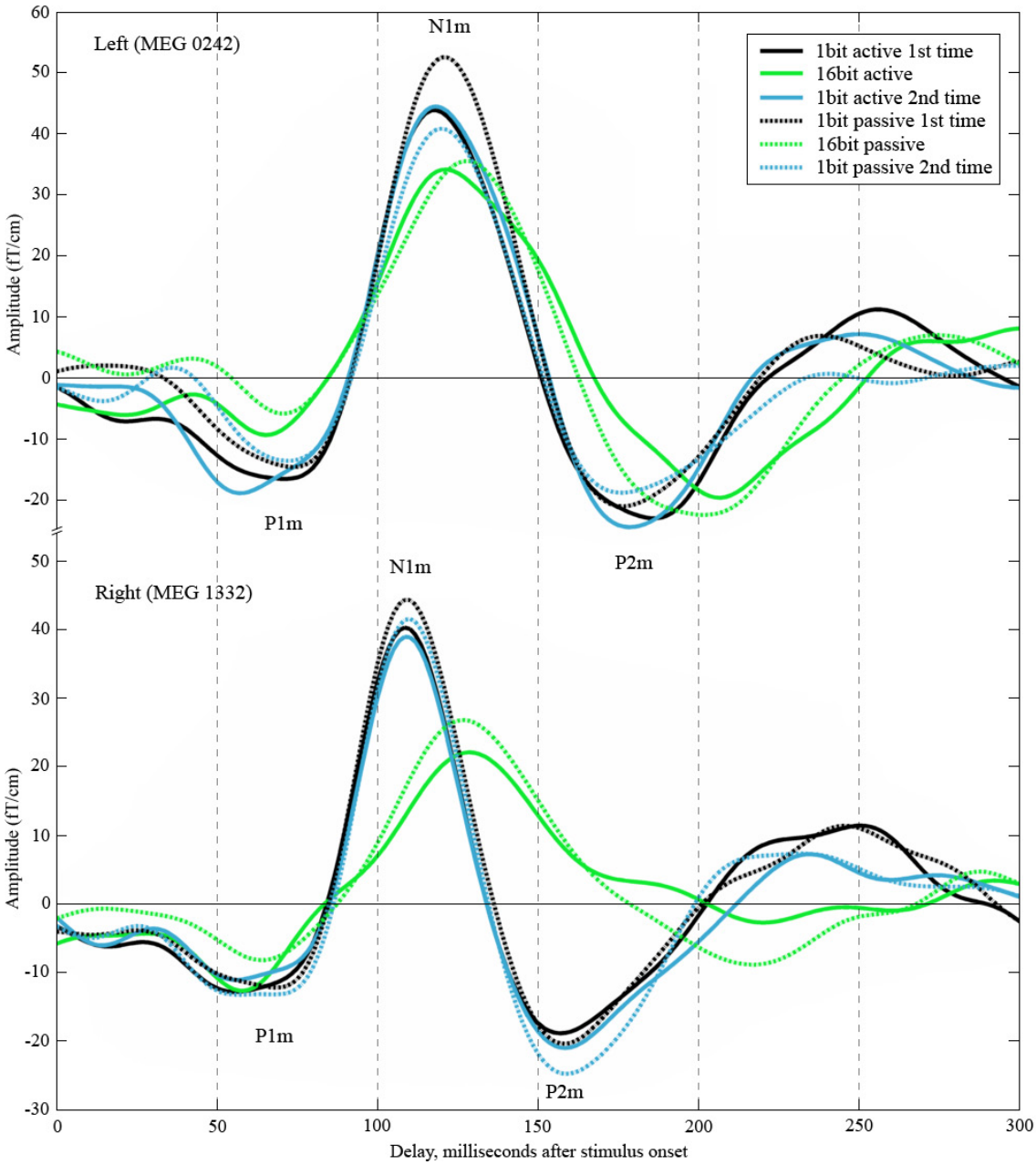


Figure 8: The grand-averaged waveforms of the transient responses in the different conditions for the left- and right-hemispheric gradiometer channels exhibiting the most prominent N1m responses (MEG 0242 and MEG 1332). The solid lines represent the response to the sentences during active and the dotted lines during passive listening. The amplitudes of the P1m, N1m and P2m increased and the latencies of the N1m and P2m dramatically decreased as an effect of stimulus degradation (green vs. dark lines). The latencies were affected more in the right hemisphere (green lines, the upper vs. the lower part of the figure). Also notice the difference in amplitude between the right-hemispheric P2m responses during the first and the second presentation of the degraded sentences (black vs. blue lines). Before averaging, the epochs were baseline corrected (100 ms pre-stimulus) and band-pass filtered (2-30Hz).

The effects of sound degradation were observable equally in both hemispheres, all transient responses showing significantly higher mean amplitudes for the degraded sentences than the good quality sentences. The mean amplitude of the P1m (aggregated over the passive and active condition) rose on average from 16.3 fT/cm in the case of good-quality sentences to 23.4 fT/cm for the degraded sentences ($F[2,16] = 10.62$, $p < 0.005$). The corresponding values for the N1m were 42.8 fT/cm and 63.6 fT/cm ($F[2, 16] = 9.63$, $p < 0.005$), and for the P2m they were 23.6 fT/cm and 34.1 fT/cm ($F[2, 16] = 5.45$, $p < 0.05$). No statistically significant hemispherical asymmetries related to sound degradation were observed in the response amplitudes. In addition to the effects of stimuli degradation, interestingly, the right-hemispheric P2m response was stronger during the second than the first presentation of the degraded sentences ($F[2,16] = 8.04$, $p < 0.05$). During the first presentation, the amplitude of the right-hemispheric P2m response was on average 34.6 fT/cm, whereas during the second presentation the mean amplitude was 39.6 fT/cm. Attention was not found to have any effects on the amplitudes of the transient responses.

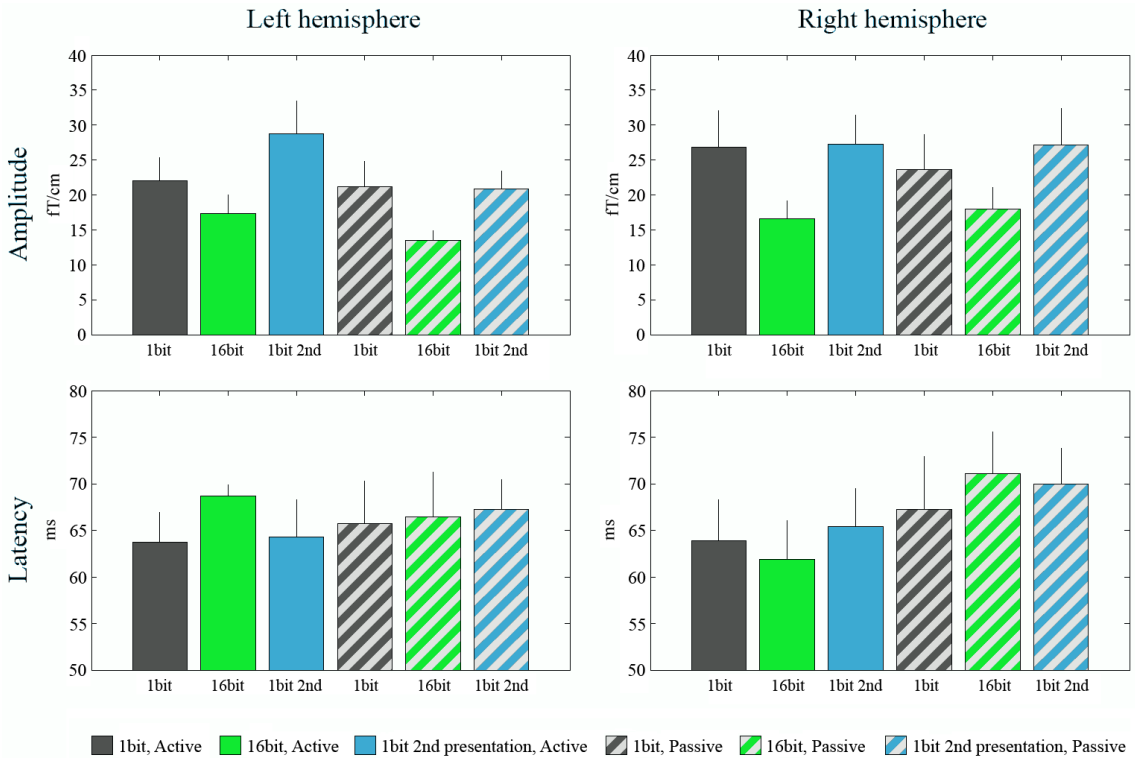


Figure 9: Mean amplitudes and latencies of the P1m response. The solid bars indicate mean amplitudes and latencies during active listening whereas the striped bars show values during passive condition. The amplitude of the P1m increased as an effect of bit degradation (dark vs. green bars). No latency effects were observed in the case of P1m. Error bars indicate SEM.

The latencies of the N1m and P2m responses were affected by stimulus degradation in both hemispheres. The latencies of the responses were significantly lower for

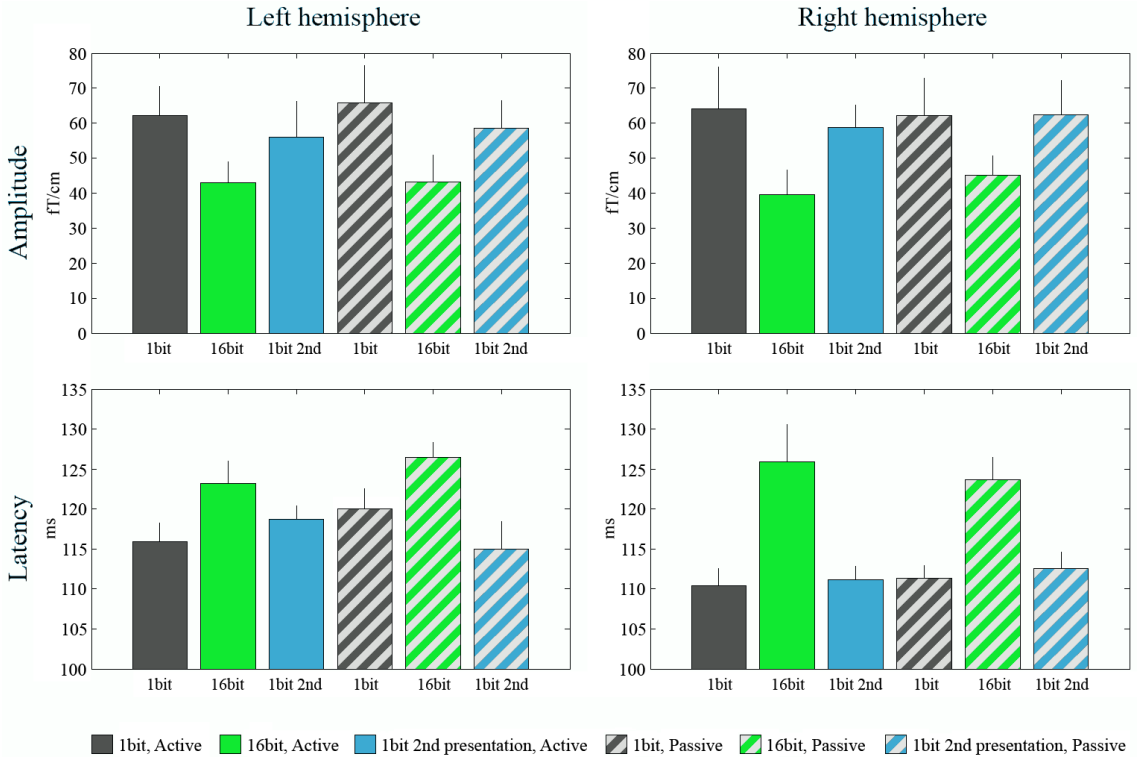


Figure 10: Mean amplitudes and latencies of the N1m response. The degraded sentences resulted in larger amplitudes and shorter latencies than the good-quality sentences (dark vs. green bars). The latency of the N1m decreased more dramatically in the right than in the left hemisphere (bottom-right vs. bottom-left figure). Error bars indicate SEM.

the degraded 1-bit sentences than for the good quality 16-bit sentences, the mean latency of the N1m decreasing from 124.8 ms to 114.4 ms ($F[2, 16] = 19.07$, $p < 0.0001$) and that of P2m from 201.7 ms to 172.1 ms ($F[2, 16] = 17.91$, $p < 0.0001$). Separate comparison revealed that the latencies of the N1m and the P2m in the left and right hemispheres were differently affected by stimulus degradation ($F[1, 8] = 10.68$, $p < 0.05$). The right-hemispheric latencies of N1m were more dramatically affected by the degradation than the left-hemispheric ones. The latency of the N1m in the left hemisphere decreased due to stimulus degradation on average by 6.9 ms (from 124.8 to 118.0 ms), whereas in the right hemisphere, the mean latency fell by 14.0 ms (from 124.8 to 110.8 ms). On average, the N1m responses peaked later in the right hemisphere for all of the stimuli types ($F[1, 8] = 11.47$, $p < 0.01$). In the case of the P2m, the difference between the hemispheres was even larger, the left-hemispheric mean latency decreasing by 20.4 ms (from 198.8 to 178.4 ms) and the mean latency in the right hemisphere dropping by 38.8 ms due to stimulus degradation (from 204.6 to 165.8 ms); ($F[2,16] = 7.04$, $p < 0.01$). All of the effects on latency were present during both active and passive listening conditions.

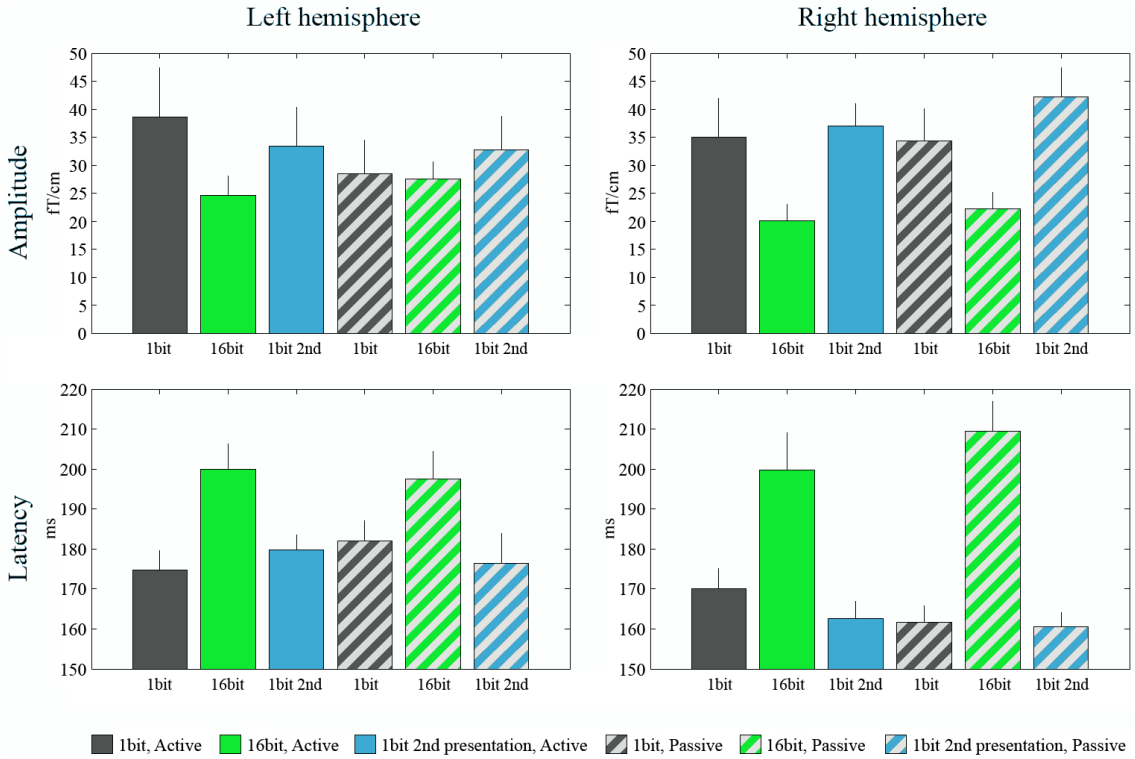


Figure 11: Mean amplitudes and latencies of the P2m response. The P2m displayed similar effects of stimulus degradation as the P1m and the N1m (dark vs. green bars), the right-hemispheric latency decreasing more by stimulus degradation than the left-hemispheric latency (notice the difference between bottom-right and bottom-left figures). Interestingly, the right-hemispheric amplitude of P2m was higher during the second presentation of the degraded sentences (top-right figure, blue vs. dark bars). Error bars indicate SEM.

### 4.2.2 Source generators of the transient responses

The source locations of the P1m, N1m and the P2m were studied with ECD modeling. Source localization could be reliably performed for all of the responses, the goodness-of-fit value ($g$) for the ECDs being on average 0.87. In both left and right hemisphere, the sources of the N1m were localized to the superior temporal region, the respective mean ECD $x$, $y$, and $z$ coordinates being -50.7, 1.1, 61.7 mm and 52.6, 9.6, 63.2 mm. In the case of the P1m, the corresponding values were -49.3, 0.6, 61.2 mm and 51.0, 5.3,63.8 mm and in the case of the P2m, they were -46.0, 1.9, 63.8 mm and 47.6, 11.3, 61.3 mm. The estimated right-hemispheric generators were found, on average, significantly more anterior during both the N1m ($F[1, 8] = 22.41$, $p < 0.005$) and P2m ($F[1, 8] = 10.92$, $p < 0.05$) responses. The difference between the left- and right-hemispheric mean $y$-coordinates was for the N1m 8.5 mm and for the P2m 9.4 mm. Estimated mean locations of the sources for the transient deflections in the different conditions are presented in Figures 12, 13 and 14.

There were considerable source location differences between the responses elicited

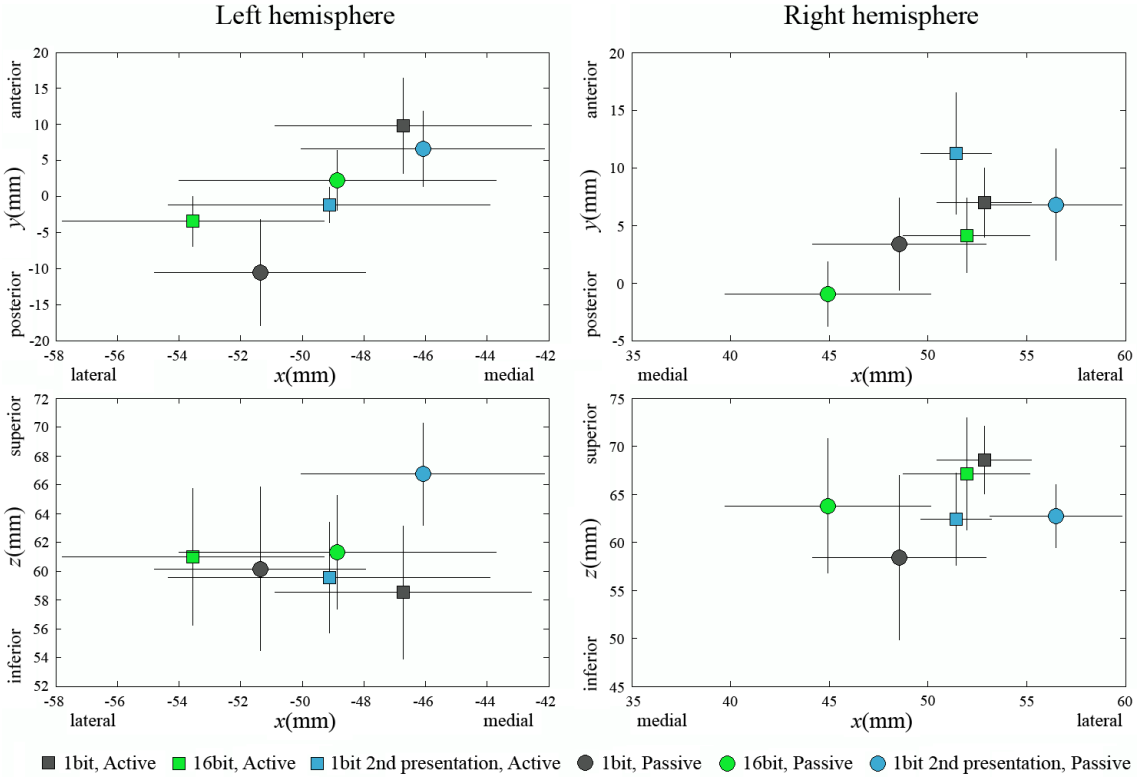


Figure 12: Mean ECD locations for the P1m response. No effects were found significant in the case of P1m ECDs. Error bars indicate SEM.

by the good-quality and the degraded sentences. In the left hemisphere, the ECDs for the N1m response elicited by the degraded sentences, were significantly more medial compared to those in the case of the good-quality sentences ($F[2,16] = 4.57$, $p < 0.05$). The $x$-coordinates of the ECDs for the degraded sentences were 4.6 mm more medial, having a value of -54.0 mm for the 16-bit and -49.5 mm for the 1-bit-sentences. In the case of the P2m, stimulus degradation resulted in a larger and more robust shift of location, the response for the degraded sentences originating from a more posterior source generator than that for the good-quality sentences ($F[2, 16] = 18.89$, $p < 0.0001$). The effect was present on both hemispheres, the mean $y$-coordinate of ECDs moving a distance of 9.2 mm (from -1.2 mm to 8.0 mm) in the left and 8.0 mm (from 8.0 mm to 19.0 mm) in the right hemisphere. There were no statistically significant differences between the source locations for the responses elicited during the first and the second presentation of the degraded sentences.

Source localization revealed effects of attention on the N1m and P2m responses. Compared to the results from the passive condition, the locations of the ECDs during N1m were more superior when the subjects were actively attending to the stimuli ($F[1, 8] = 6.54$, $p < 0.05$). The difference between the mean $z$-coordinates of the ECDs during passive and active listening was in the left hemisphere on average 5.7 mm (58.9 mm and 64.6 mm) and in the right hemisphere 1.3 mm (62.6 mm and 63.8

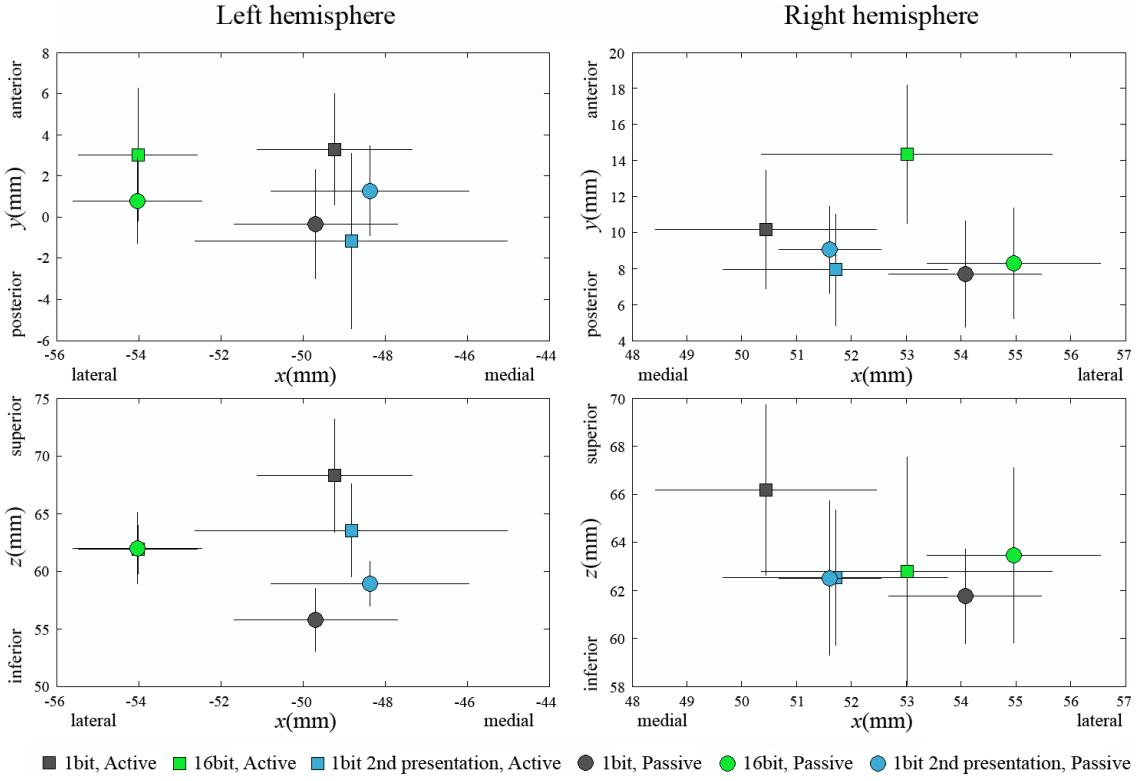


Figure 13: Mean ECD locations for the N1m. The left-hemispheric ECDs were more medial for the degraded sentences (dark vs. green markers). In addition, the N1m ECDs were shifted in the superior direction during the active listening condition (squares vs. circles). The right-hemispheric N1m ECDs were under all conditions, more anterior compared to the left-hemispheric ones (top-right vs. top-left figure). Error bars indicate SEM.

mm). In the case of P2m, during active listening condition, the right-hemispheric sources were located more medially, the difference between the mean $x$-coordinates being approximately 7.1 mm (active: 44.0 mm, passive: 51.1 mm, $F[1, 8] = 9.01$, $p < 0.05$).

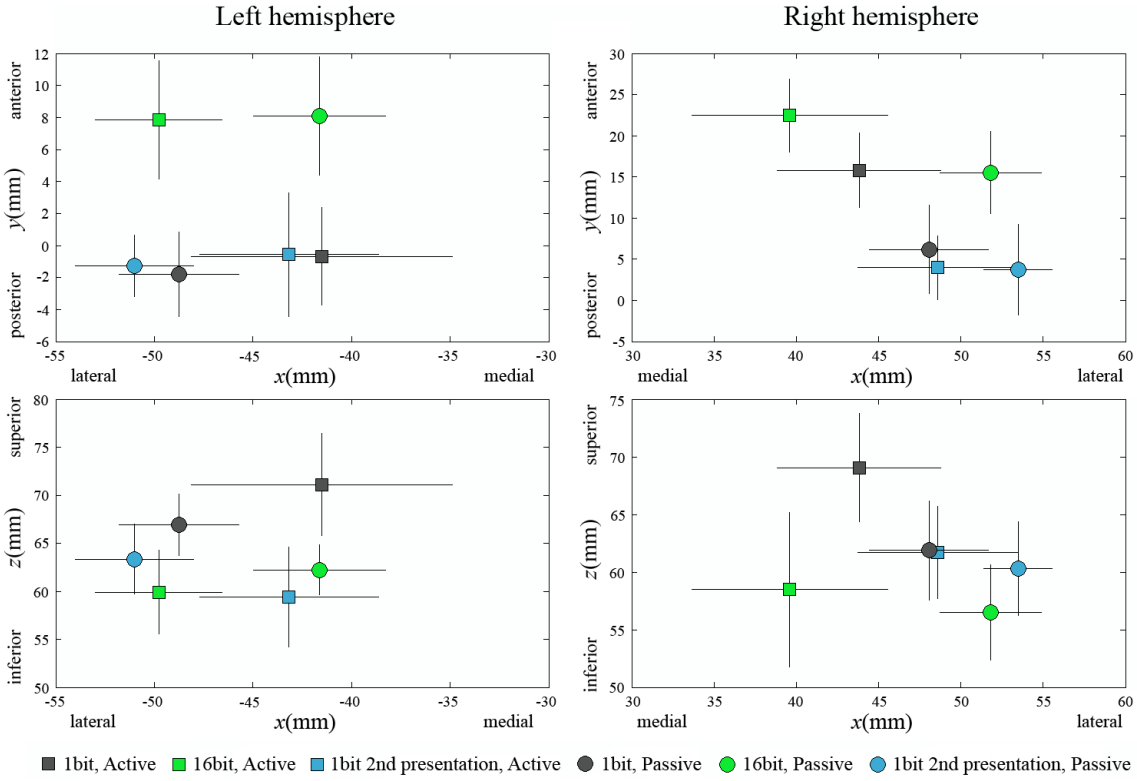


Figure 14: Mean ECD locations for the P2m. The left-hemispheric sources of the P2m response were more posterior when a degraded sentence was presented (dark vs. green markers of the upper-left figure). In the right hemisphere, the P2m ECDs were located more medial during active listening (right figures, squares vs. circles). Similarly to the N1m, in general, the sources of the P2m were more anterior in the right than in the left hemisphere (right vs. left figures). Error bars indicate SEM.

### 4.2.3 Mean amplitude of the sustained field

Two clearly separate time intervals were distinguished in the SF. The low-pass filtered waveforms of the grand-averaged AEFs in the left and right hemisphere are shown in Figure 16. The SF began as a considerable negative deflection that started to build up at approximately 300 ms, peaked between 500 - 700 ms and tapered off at around 1000 ms after stimulus onset. After this, for the remainder of the stimulus duration, the amplitude of the SF remained relatively constant. In the following, the terms early SF and late SF refer to these parts, defined as the AEF during the intervals of 300 - 1000 ms and 1000 - 3000 ms. To study the early and late SF, mean amplitudes were calculated over the two intervals and are are depicted in Figure 15.

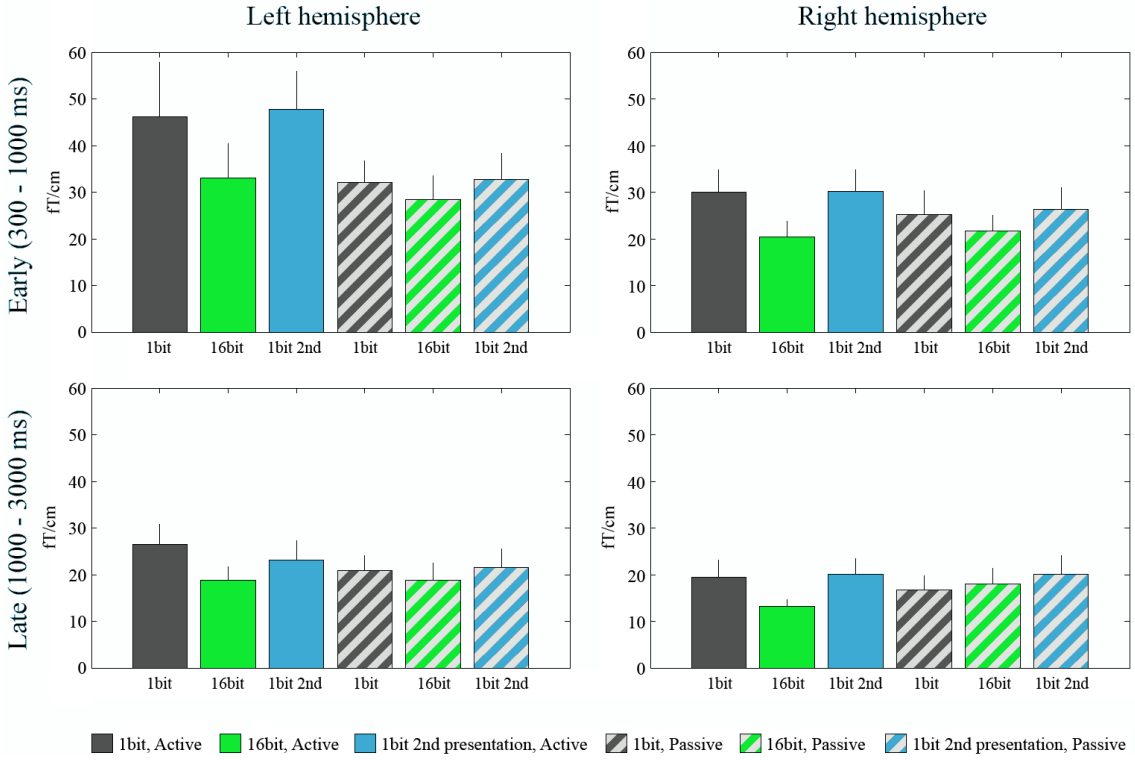


Figure 15: Grand-averaged mean amplitudes during the early (300 - 1000 ms) and late (1000 - 3000 ms) SF in the left and right hemisphere. The early part had a higher mean amplitude during the active listening condition (upper figures, solid vs. striped bars). Overall, the early part had a higher amplitude than the late part (upper vs. lower figures).

The early SF, was confirmed to have a significantly higher mean amplitude than the late SF ($F[1, 7] = 19.20$, $p < 0.005$). The mean amplitude during the early SF ranged from 20.4 fT/cm to 47.8 fT/cm, while for the late SF, the mean amplitude ranged from 13.3 fT/cm to 26.4 fT/cm.

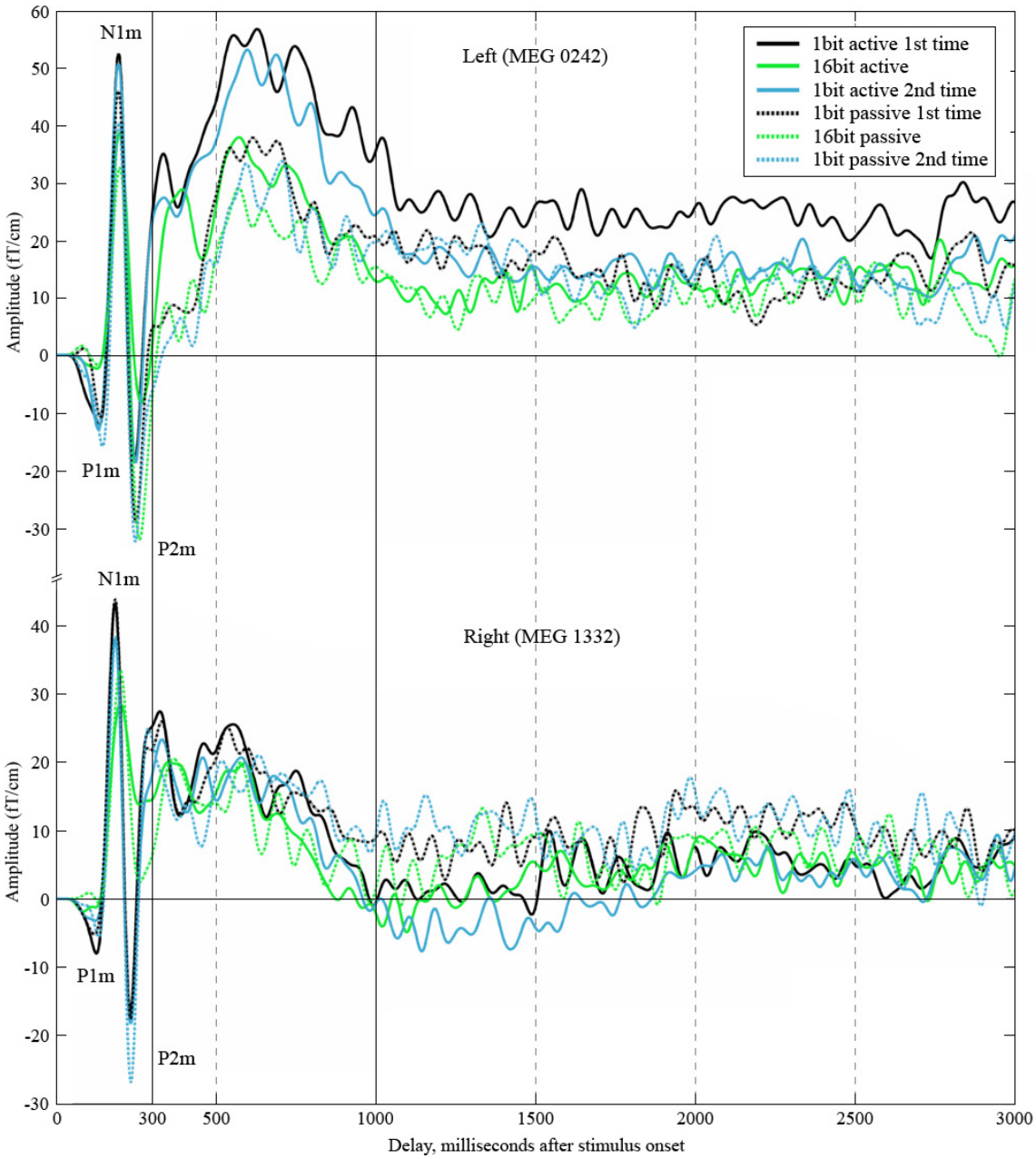


Figure 16: The grand-averaged and low-pass filtered AEF from the left- and right-hemispheric gradiometer channels displaying the maximum N1m response (MEG 0242 and MEG 1332). For visualization, the signals were low-pass filtered using a cutoff frequency of 15Hz. The solid vertical lines separate the time intervals used to calculate the mean amplitudes of the SF. The sentences elicited more prominent responses during 300-1000 ms when the subject was actively attending to the stimuli (solid vs. dotted lines). Additionally, the early SF had a higher amplitude during the presentation of the degraded sentences (dark vs. green line) and was stronger in the left hemisphere (upper vs. lower part of the figure).

The early SF had a higher mean amplitude in the active listening condition than in the passive condition ($F[1, 7] = 7.81$, $p < 0.05$). During the active condition, the mean amplitude of the early SF was on average 34.6 fT/cm, whereas during the passive condition, the mean amplitude was only 27.7 fT/cm. The grand-averaged amplitudes for the early SF had a similar pattern as the amplitudes of the transient responses, showing higher values for the degraded sentences. The mean amplitude of the early SF elicited by the degraded and good-quality sentences was 33.3 fT/cm and 25.8 fT/cm, respectively. The effect approached statistical significance, the significance level reaching $p = 0.052$ ($F[1, 7] = 3.67$). Comparison between the left- and right-hemispheric amplitudes of the early SF revealed signs of hemispheric asymmetry, the left-hemispheric mean amplitudes being larger. The left- and right-hemispheric amplitudes aggregated over all of the conditions were 36.7 fT/cm and 25.6 fT/cm respectively. The effect approached statistical significance ($F[1, 7] = 5.05$, $p = 0.059$).

In the case of the late SF, no effects of attention, stimulus degradation or hemisphere were found. Additionally, there was no notable difference between the SF elicited during the first and the second presentation of the degraded sentences.

### 4.2.4 Current distribution during the transient responses

To study the current distributions, MNEs were computed for the P1m, N1m and the P2m responses by extracting a 40-ms time window around the response peaks. The received information on the average current spread over the cortex was further analyzed by dividing the cortical surface into 24 regions and subsequently calculating the mean currents over the voxels inside these regions (ROI analysis). Additionally, for visualization of the responses, noise-normalized MNEs were calculated by delineating the MNEs using sLORETA estimates.

The P1m response exhibited a widespread current distribution, the most activated areas being difficult to localize visually (see Figure 18). However, ROI analysis revealed numerous regions affected by stimulus degradation, attention and learning (see Figure 17). Here, effects of learning were defined as differences between the first and the second presentation of the degraded sentences. In both hemispheres the degraded sentences elicited higher currents in the central superior temporal region (CST) than the good-quality sentences. For the 16-bit-sentences the mean current in the CST was 4.8 pA/m whereas for the 1-bit-sentences the current was on average 6.1 pA/m ($F[2, 14] = 6.68$, $p < 0.01$). Attention modulated the central inferior temporal regions (CIT) in both hemispheres and the right-hemispheric posterior superior parietal region (PSP), the areas showing higher currents in the passive than in the active listening condition. During the active condition the current for the bihemispheric CIT was on average 4.6 pA/m, whereas during the passive condition, it was 5.7 pA/m ($F[1, 7] = 6.01$, $p < 0.05$). The right-hemispheric PSP behaved similarly, the corresponding values being 8.0 pA/m and 10.4 pA/m ($F[1, 7] = 6.32$, $p < 0.05$). Interestingly, the P1m was found to be sensitive to learning. The an-

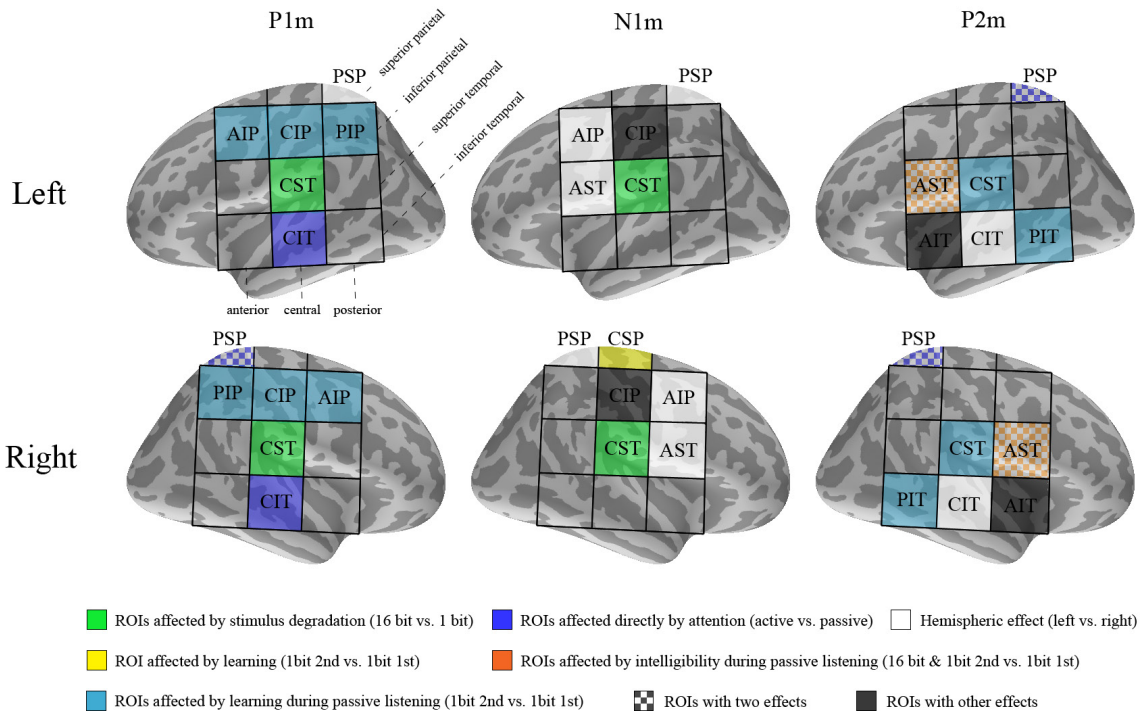


Figure 17: Regions of interest affected by stimulus quality, learning and attention during P1m, N1m and P2m. Figures show the statistically significant effects ($p < 0.05$) as revealed by ANOVA.

terior inferior parietal region (AIP, $F[2, 14] = 4.27$, $p < 0.05$), the central inferior parietal regions (CIP, $F[2, 14] = 8.14$, $p < 0.005$), and the posterior inferior parietal regions (PIP, $F[2, 14] = 3.75$, $p < 0.005$) of both hemispheres showed interactions of attention and bit mode. Post-hoc tests revealed that activity within the AIP (post-hoc test $p < 0.05$) and the CIP (post-hoc test $p < 0.005$) was significantly stronger during the second presentation of the degraded sentences of the passive listening condition than during any of the other five conditions. During this condition, the mean current was for the AIP 9.3 pA/m and for the CIP 9.0 pA/m whereas during all of the other conditions, the values within the two regions were in the range of 5.3 pA/m and 6.7 pA/m. The interaction within the PIP was slightly different, the second presentation of the 1-bit-sentences during passive condition eliciting larger mean currents than any except the 16-bit-sentences during the passive condition (post-hoc test $p < 0.05$). Current within the PIP during this condition was 9.2 pA/m while in the other conditions it ranged from 5.7 pA/m to 6.5 pA/m. Lastly, a hemispheric effect was found within the PSP, the right-hemispheric region showing higher amplitudes (9.2 pA/m) than the left-hemispheric one (7.2 pA/m, $F[1, 7] = 8.67$, $p < 0.05$).

During the N1m, the source current was clearly localized to the vicinity of the auditory cortices, the maximum current being under all studied conditions located in the superior temporal region (see Figure 19). Similarly to the single-sensor measurements, the responses were evidently stronger during the degraded stimuli than

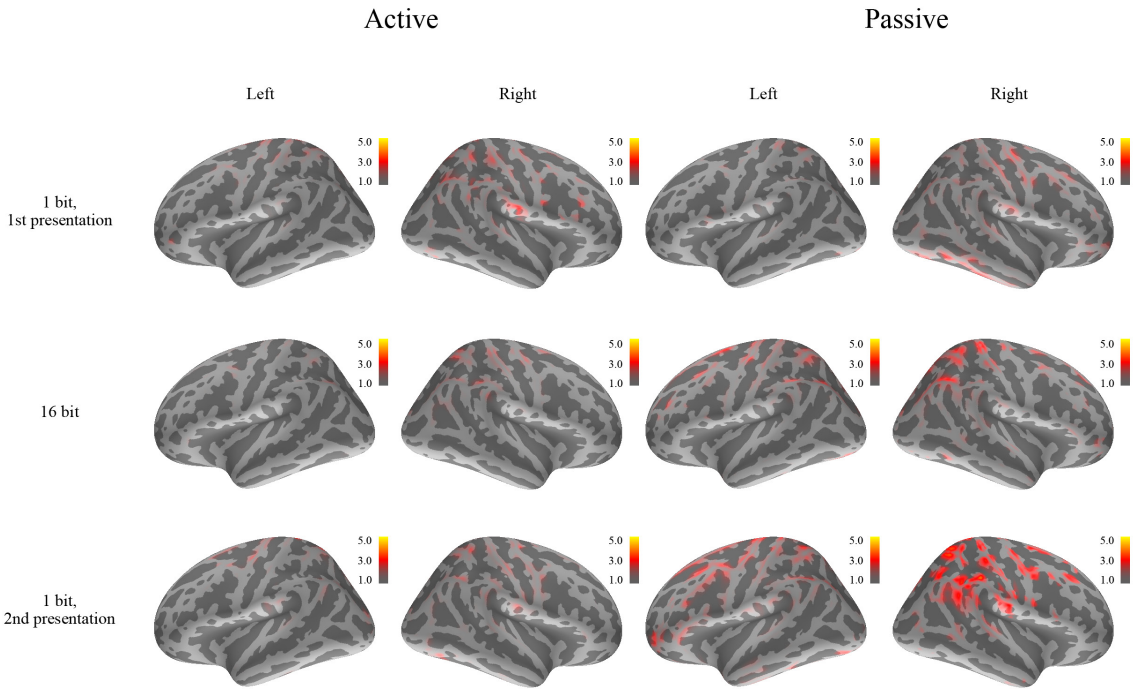


Figure 18: Noise-normalized current distribution during the P1m visualized on the surface of an average brain. The noise-normalized MNE was averaged in a ±20 ms time window around the peak of the P1m response (threshold = 50 % of sLORETA estimates maximum value). Note especially the difference between the current distributions within the inferior parietal regions elicited during the first and the second presentation of the degraded sentences in the passive condition ($3^{rd}$ and $4^{th}$ column, top vs. bottom figures). Also, observe the increased activation within the right-hemispheric PSP in the passive condition ($4^{th}$ column vs. $2^{nd}$ column).

during the good-quality stimuli. These effects were confirmed with ROI analysis. The CST had a significantly higher mean current during the presentation of the 1-bit-sentences than the 16-bit-sentences ($F[2, 14] = 5.62$, $p < 0.05$). The mean currents during the 1-bit- and 16-bit-sentences within the CST were 15.0 pA/m and 11.5 pA/m, respectively. The CIP showed a main effect of bit mode ($F[2, 14] = 4.17$). Post-hoc test revealed that current within the CIP was higher during the second presentation of the degraded sentences than during the good-quality sentences, the value being 12.9 pA/m and 10.0 pA/m, respectively (post-hoc test $p < 0.05$). Also the N1m was found to reflect learning, the right-hemispheric central superior parietal region (CSP) exhibiting stronger activation during the second presentation of the degraded sentences than during the good-quality sentences and the first presentation of the good-quality sentences (separate test for the right-hemispheric CSP, $F[2, 14] = 4.71$). This effect was observable in both the active and the passive condition. During the second presentation of the 1-bit-sentences, the right-hemispheric CSP had a mean current of 12.4 pA/m while during the first presentation of 1-bit-sentences it was 8.3 pA/m and during the 16-bit-sentences it was 9.1 pA/m (post-hoc test $p < 0.05$). Finally, a hemispheric effect was found in the anterior superior tem-

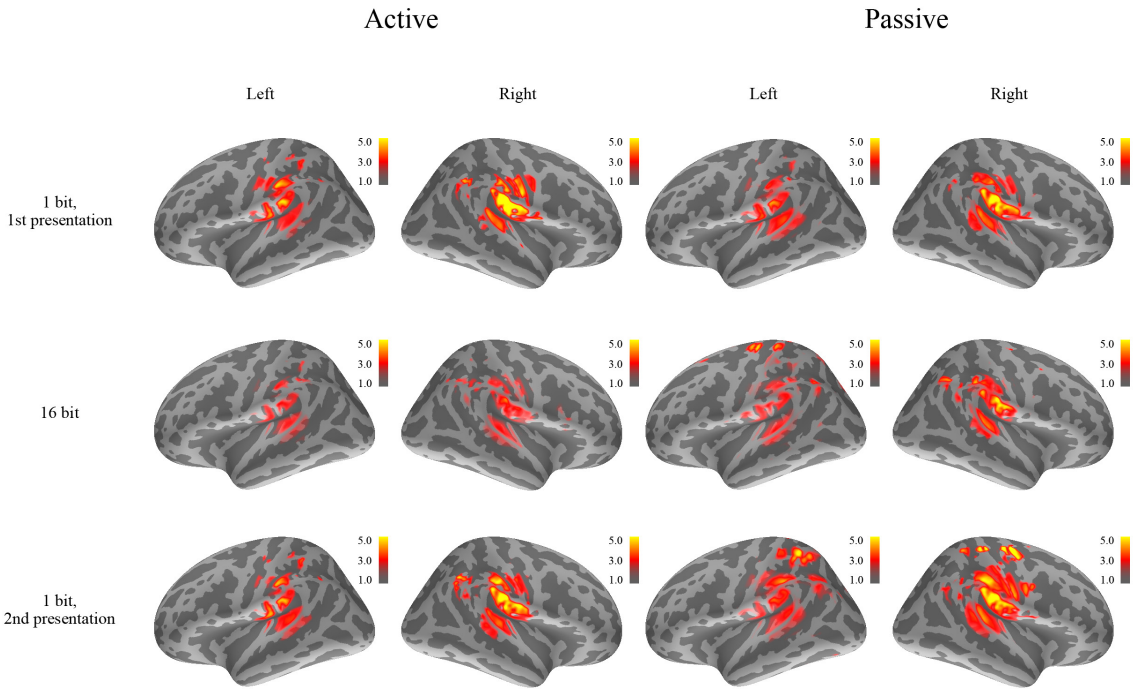


Figure 19: Noise-normalized current distribution during the N1m. The N1m source current was clearly localized to the vicinity of the auditory cortices. Note the smaller currents during the good-quality sentences (second row vs. first and third rows). Also, note the increased activation within the right-hemispheric CSP during the second presentation of the degraded sentences ($4^{\text{th}}$ column, $3^{\text{rd}}$ vs. $1^{\text{st}}$ and $2^{\text{nd}}$ figure)

poral region (AST, $F[1, 7] = 9.83$, $p < 0.05$), in the AIP ($F[1, 7] = 19.81$, $p < 0.005$), and in the PSP ($F[1, 7] = 7.78$, $p < 0.05$), the right-hemispheric areas having higher mean current than the left hemispheric ones. The left- and right-hemispheric mean currents were for the AIP 10.6 pA/m and 7.4 pA/m, for the AST 11.3 pA/m and 7.3 pA/m and for the PSP 11.2 pA/m and 9.3 pA/m, respectively. Attention had no direct effect on the current distributions during the N1m.

The P2m did not produce as focal current estimates as the N1m response. The estimated source current was spread within a larger radius, indicating a non point-like source (see Figure 20). Similarly to the P1m, also the P2m showed effects of learning during the passive listening condition. Four regions exhibited interaction of attention and bit mode: the AST ($F[2, 14] = 6.01$), the posterior inferior temporal region (PIT, $F[2, 14] = 4.61$), the CST ($F[2, 14] = 5.48$), and the AIT ($F[2, 14] = 4.77$). Post-hoc tests showed that these interactions were all slightly different. In the case of the AST, in the passive condition, the good-quality sentences and the second presentation of the degraded sentences elicited higher currents than the first presentation of the degraded sentences (post-hoc test $p < 0.05$). Mean current within the AST in the passive condition was for the the first presentation of 1-bit-sentences 7.7 pA/m, for the 16-bit-sentences 10.5 pA/m, and for the second presentation of

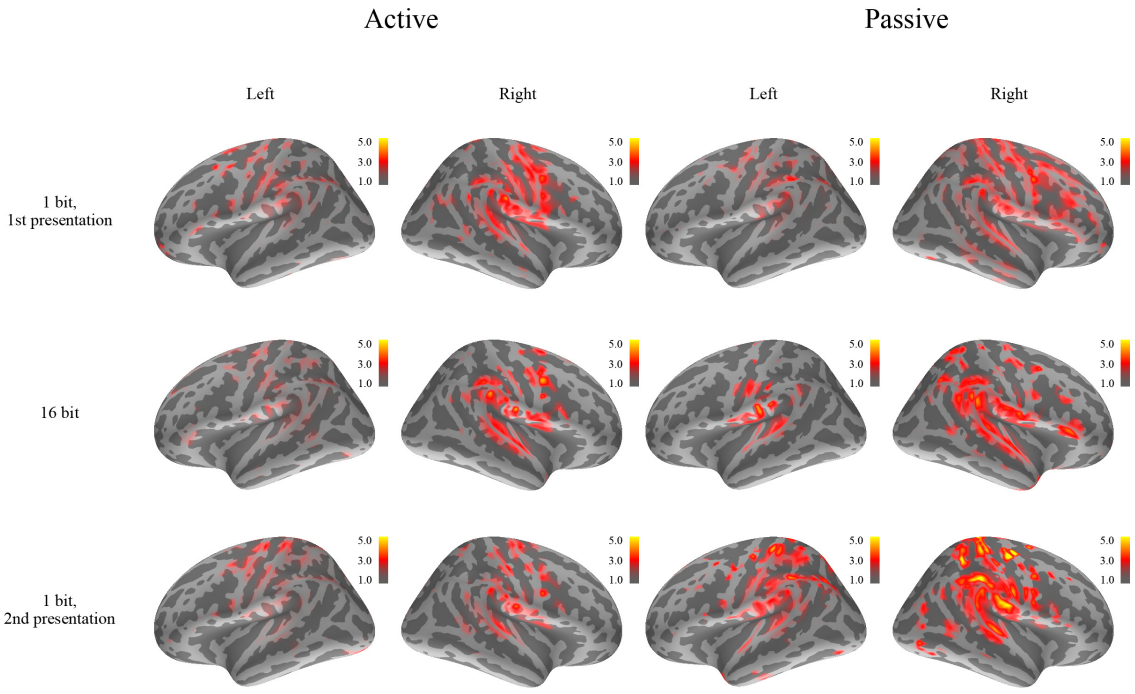


Figure 20: Noise-normalized current distribution during the P2m. The P2m source current was spread within a large radius, indicating a non point-like source. Note the increased activation within the temporal region during the second presentation of the degraded sentences in the passive condition ($3^{rd}$ and $4^{th}$ column, bottom vs. top and middle figures). Also, observe the difference between the activation within the PSP during the passive and active condition ($3^{rd}$ and $4^{th}$ vs. $1^{st}$ and $2^{nd}$ column).

1-bit-sentences 10.6 pA/m. Within the PIT, the currents elicited during the second presentation of the degraded sentences in the passive condition were higher than those elicited during any other presentation of degraded sentences (post-hoc test $p < 0.05$). The mean current was for this condition 9.2 pA/m, whereas during the first and second presentation of degraded sentences in the active condition and the first presentation of degraded sentences in the passive condition, the value ranged from 6.6 pA/m to 7.1 pA/m. In the case of the CST, the second presentation of 1-bit-sentences in the passive condition elicited higher currents than the first presentation of the 1-bit-sentences in the passive and the second presentation of the 1-bit-sentences in the active conditions (post-hoc test $p < 0.05$). The mean current was for this condition 12.5 pA/m, whereas for the first presentation of degraded sentences in the passive condition and the second presentation in the active condition, the mean currents were 8.8 pA/m and 9.0 pA/m, respectively. Within the AIT, the activation elicited by the good-quality sentences in the passive condition (7.1 pA/m) was higher than that elicited by the first presentation of the degraded sentences in the passive condition (5.3 pA/m) and the second presentation of the degraded sentences in the active condition (5.2 pA/m, post-hoc test $p < 0.05$). The P2m was modulated by attention. The left- and right-hemispheric PSP had

a higher mean current during the passive condition, the values being 9.23 pA/m and 11.2 pA/m for the active and passive conditions respectively ($F[1, 7] = 5.76$, $p < 0.05$). Lastly, a hemispheric effect was found in the AST ($F[1, 7] = 7.78$, $p < 0.05$), in the CIT ($F[1, 7] = 6.44$, $p < 0.05$) and in the PSP ($F[1, 7] = 5.76$, $p < 0.05$), the right-hemispheric areas being overall more strongly activated. The left- and right-hemispheric mean currents, aggregated over the conditions, were for the AST 7.4 pA/m and 10.9 pA/m, for the CIT 7.1 pA/m and 9.9 pA/m and for the PSP 11.6 pA/m and 8.9 pA/m, respectively.

### 4.2.5 Current distribution during the sustained response

The noise-normalized MNEs showed that the sustained response has a widespread character, the activation extending to temporal and parietal areas. It is fairly difficult to identify distinct sources. However, comparing the early and the late SF, the same areas are activated during the periods, but the strength of the currents is lower during the late sustained response. The mean current distributions for the early and the late SF over the studied factors are shown in Figures 22 and 23, respectively.

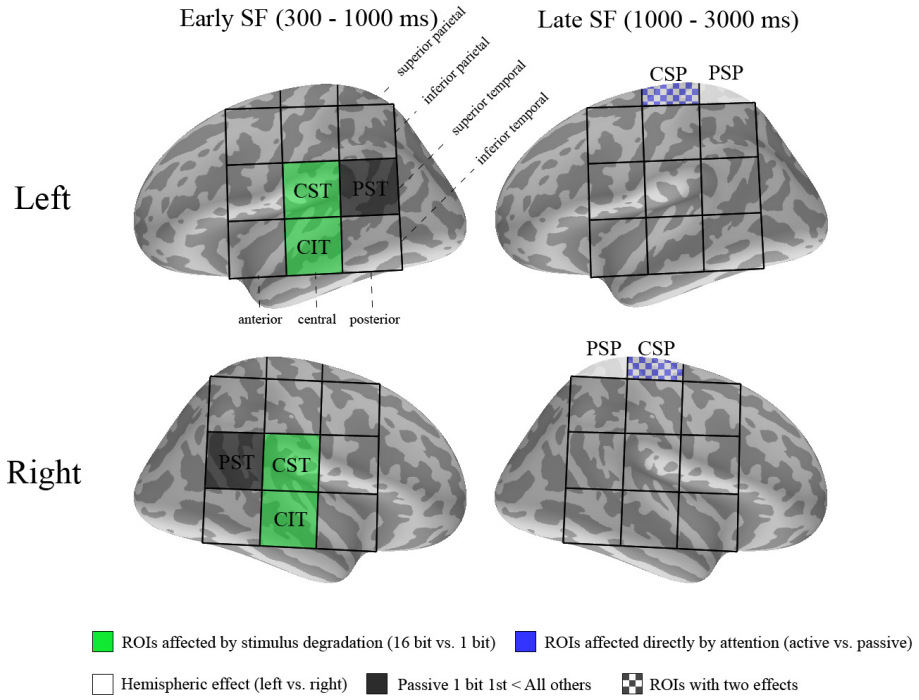


Figure 21: Regions of interest affected by the stimulus quality, learning and attention during the early (300 - 1000 ms) and the late (1000 - 3000 ms) SF. Figures show the statistically significant effects ($p < 0.05$) as revealed by ANOVA.

The effects on the early and late SF within the ROIs are presented in Figure 21. ROI analysis confirmed that the early SF is affected by sound degradation. The mean currents within the CST and the CIT were higher when a degraded sentence

was presented. During the good-quality sentences, mean current within the CST was on average 8.6 pA/m, whereas during the degraded sentences it was 11.3 pA/m ($F[2, 14] = 5.28$, $p < 0.05$). In the case of the CIT, the corresponding values were 7.8 pA/m and 9.7 pA/m ($F[2, 14] = 3.89$, $p < 0.05$). In addition, an interaction between attention and stimulus degradation was found significant within the posterior superior temporal region (PST, $F[2, 14] = 4.20$). Post-hoc test revealed that activation within the PST, elicited during the first presentation of the degraded stimuli in the passive listening condition, was significantly lower than that elicited during any of the other conditions (post-hoc test $p < 0.05$). During this session, the elicited mean currents within the PST were on average 7.8 pA/m whereas during the other sessions, the mean currents ranged from 9.3 pA/m to 10.2 pA/m.

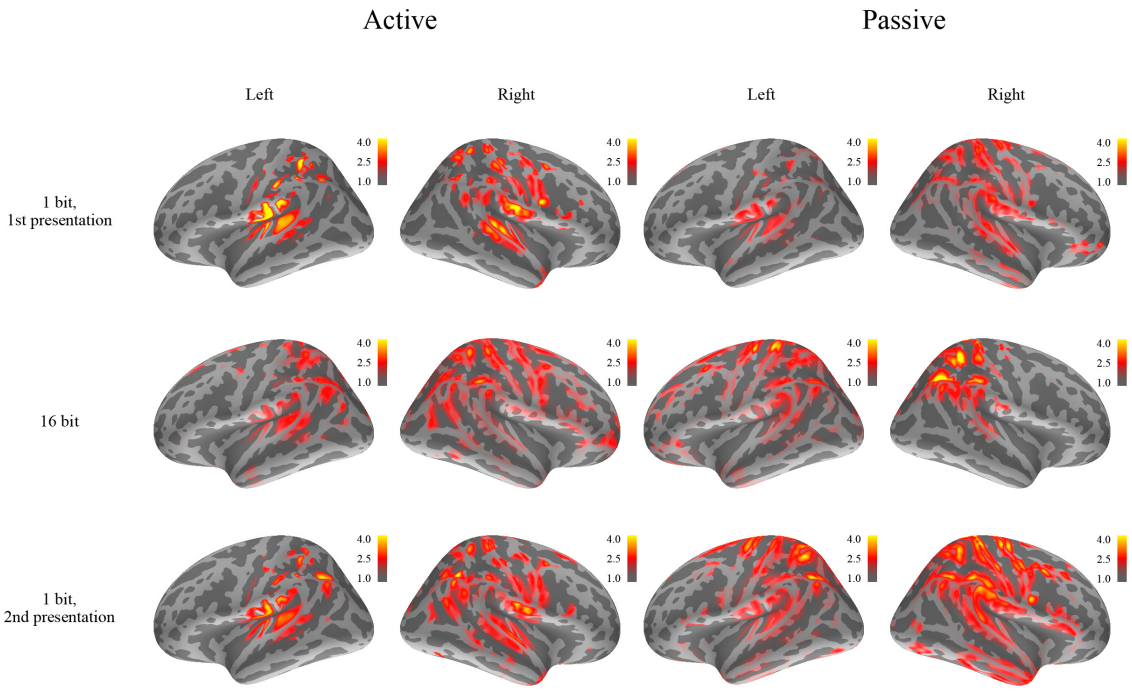


Figure 22: Current distributions during the early SF (300 - 1000ms). During the early SF, stimulus degradation resulted in increased activation within the CST and CIT ($1^{st}$ vs. $2^{nd}$ row).

In the case of the late SF, two effects were found in the superior parts of the parietal lobe. The late SF was found to be sensitive to attention. The left- and right-hemispheric CSP exhibited higher amplitudes during the passive condition (10.5 pA/m) than during the active condition (8.4 pA/m, $F[1, 7] = 6.84$, $p < 0.05$). Also, a hemispheric effect was found within the CSP ($F[1,7] = 10.05$, $p < 0.05$) and the PSP ($F[1,7] = 11.33$, $p < 0.05$), the right-hemispheric regions overall having higher mean currents than the left-hemispheric ones. The left- and right-hemispheric mean currents were for the CSP 8.8 pA/m and 10.1 pA/m and for the PSP 9.7 pA/m and 12.0 pA/m, respectively.

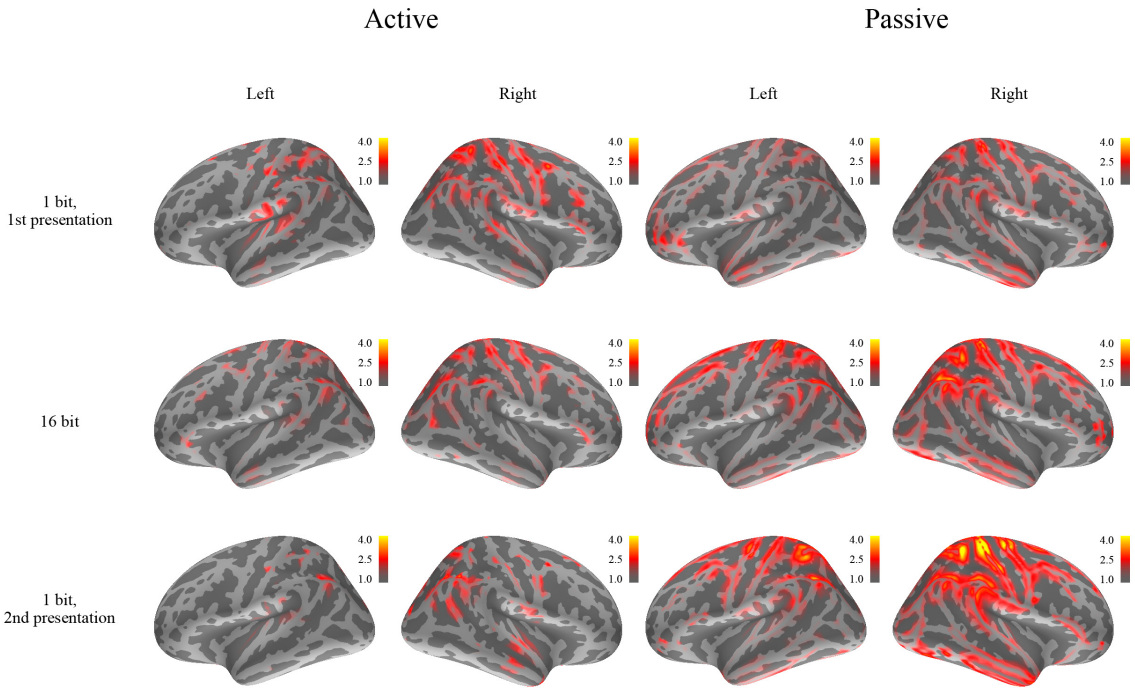


Figure 23: Current distribution during the late SF (1000 - 3000 ms). In most of the conditions, the extent of the activation appears to be similar to the activation during the early SF (300 - 1000 ms). Note the increased activation within the CSP during the passive condition ($3^{rd}$ and $4^{th}$ vs. $1^{st}$ and $2^{nd}$ column).

# 5 Summary and discussion

The present study aimed to study three topics: the cortical mechanisms underlying our capability to understand speech (1) under difficult circumstances, (2) under varying levels of *a priori* knowledge, and (3) under varying levels of attention. These goals were achieved by taking advantage of the utterance conveyed by words in complete Finnish sentences and the controlled reduction of sound quality provided by the USQ method. The perceptual quality of the sentences was degraded to a level where comprehension was initially difficult. However, if the degraded sentences were preceded by the original sentences, the degraded sentences became comprehensible by the subject. Thus, importantly, during the first and the second presentation of the degraded sentences, the stimuli were physically identical but only the latter one was comprehensible.

Presenting the original and the degraded sentences subsequently made it possible to observe the effects of sound degradation (1). The effects of *a priori* information on the prospective sentences (2) were studied by comparing the responses elicited during the first and the second presentation of the degraded sentences. During the second time the sentences were hypothesized to be better recognized. This hypothesis was further confirmed with both subjective and objective measures. The effects of attention (3) were studied by carrying out the procedure in active and passive condition.

Ten subjects participated to the study. Four auditory evoked magnetic responses, the P1m, N1m, P2m and the sustained response were studied using a whole-head MEG device. Information on the responses was extracted using three methods. Latencies and amplitudes of the transient response peaks were obtained directly from the calculated gradient amplitude. In the case of the SF, two intervals were studied separately: the early SF (300 - 1000 ms) and the late SF (1000 - 3000 ms). Response sources were localized using ECD modeling. Thirdly, current distribution in the cortex during the responses was estimates using MNEs and ROI analysis.

## 5.1 Effects of sound degradation

Effects of sound degradation were studied by comparing activation elicited during the first presentation of the degraded sentences to the one elicited during the presentation of the good-quality sentences. Confirming hypothesis (a) (see page 18), the degraded sounds were difficult to comprehend, only an average of 30.2 % of the sentences being understood during the first presentation. On average, 94.8 % of the good-quality sentences were understood.

Sound degradation was found to have a significant effect on all of the transient responses, amplitudes of the P1m, N1m and P2m and latencies of the N1m and P2m being altered (see Figure 8, on page 28). All of the effects were present bilat-

erally. The N1m and P2m responses elicited by the degraded sentences had higher amplitudes than those elicited by the good-quality sentences. The latencies of the N1m and P2m were significantly shorter for the degraded sentences. Despite of the fact that the effect was present in both hemispheres, the right hemisphere was more dramatically affected. As an effect of sound degradation, the latencies of the N1m and P2m in the left hemisphere decreased on average by 6.9 ms and 20.4 ms whereas in the right hemisphere, the mean latency fell by 14.0 ms and 38.8 ms. These results are congruent with those by Liikkanen et al. (2007), Miettinen et al. (2010) and by Alho (2010). Of the previous studies, only Miettinen et al. (2010) observed the amplitude effects bilaterally. However, in the present study, the effects were in some aspects more prominent, the P1m, N1m and P2m amplitudes increasing on average by 44 %, 49 % and 44 %, respectively compared to 33 %, 23 % and 32 % witnessed by Miettinen et al. (2010). Additionally, in Miettinen et al. (2010), in the case of vowel sounds, no systematic latency effects were observed.

In the case of the N1m and P2m, ECD localization of the responses revealed noticeable differences between the sources activated during the good-quality and degraded sentences (see Figures 13 and 14, p. 33 - 34). The left-hemispheric sources for the N1m were 5 mm more medial in the case of degraded sentences. The effect on P2m was more distinct, the center of gravity being bilaterally 9 mm more posterior in the case of the degraded sentences. The shift of N1m source location differs from the one in Miettinen et al. (2010), where the degraded stimuli had a location of 2 mm more anterior to the good-quality stimuli. However, it should be noted, that instead of speech, the stimuli used by Miettinen et al. (2010) comprised of vowels, sine-wave composites of vowels and pure sine-waves.

The amplitude effects on the P1m and N1m were confirmed by the ROI analysis (see Figure 17, p. 38). Importantly, whereas the single-sensor measurements accurately described the effects in the vicinity of the superior temporal regions, ROI analysis revealed regions affected throughout the cortex. In the case of the P1m, the estimated current in the central superior temporal region (CST) was higher during the presentation of the degraded sentences. As an effect of sound degradation, activation within the CST during the P1m increased on average by 27 %. The same effect was observed in the CST also during the N1m. During the N1m, the activation within the region increased on average by 30 %. In the case of P2m, no effects of stimulus degradation were observed.

The sustained response was affected by sound degradation (see Figure 15, p. 35). According to single-sensor measurements, on average, the degraded sentences elicited SF, of which early part was 29 % higher than that elicited by the good-quality sentences. Also in the case of the early SF, ROI analysis confirmed the effects observed in single-sensor data (see Figure 21, p. 42). During the early SF, activation elicited during the degraded sentences within the CST was on average 31 % higher than that elicited during the good-quality sentences. In addition to the CST, in the case of the early SF, stimulus degradation also affected central inferior

temporal region (CIT). Within the CIT, the degraded sentences elicited activation that was on average 24 % higher than that elicited by the good-quality sentences. In the case of late SF, no effect of stimulus degradation was observed.

Thus, sound degradation seems to increase activation not only within the superior, but during the late SF, also within the inferior parts of the central temporal region. This effect is present during the P1m, N1m, P2m and the early SF, the increase in amplitude being equal in both hemispheres. Degradation increases the latency of the N1m and P2m responses bilaterally, the effect being however, more prominent in the right hemisphere.

## 5.2 Effects of learning and intelligibility

In section 2.3 it was hypothesized, that listening to the comprehensible sentences prior to the degraded sentences will contribute to understanding the prospective degraded sentences. The increase of identification percentage by 48.7 percentage points between the first and the second presentation of the degraded sounds indicates a remarkable learning effect (see Figure 7, p. 27). This was further anticipated to be reflected in cortical activation, the changes representing memory-related processes. Thereby, effects of learning were studied by comparing the activations elicited during the first presentation of the degraded sentences to those elicited during the second presentation of the degraded sentences. Additionally, the effects of intelligibility were studied: whereas the first presentation a degraded sentence was, as demonstrated by the behavioral data, unintelligible, the following good-quality and degraded sentences were intelligible.

Effects of learning were reflected in the AEF early on, the current distribution of the transient P1m, N1m and P2m being each altered. In the case of the P2m, the effect was also present in the single-sensor data. Many of the found effects were interactions of attention and bit mode. Post-hoc tests revealed that these effects were mostly present only in the passive listening condition. In the case of the P1m, learning affected the whole inferior parietal region (the AIP, CIP and the PIP). Activation within the regions, in the passive condition, elicited during the second presentation of degraded sentences was higher than that elicited during the first presentation of the degraded sentences. The effect was considerable. During the second presentation of the degraded sentences in the passive condition, the mean currents within the AIP, CIP and the PIP were 48 %, 41 % and 52 % higher than during the first presentation of the degraded sentences. In the case of the N1m, learning increased the activity in the right-hemispheric central superior parietal region (CSP). Unlike during the P1m, the effect during the N1m was present during both the active and the passive listening condition. Within the CSP, compared to the activation elicited during the first presentation of the degraded sentences, the activation elicited during the second presentation of the degraded sentences was on average 49 % higher. In the case of the P2m, the peak amplitudes extracted from

the single-sensor data (see Figure 11, p. 31) showed effects of learning. During the second presentation of the degraded sentences, the P2m amplitude was 14 % higher than during the first presentation of the same sentences. ROI analysis confirmed this (see Figure 17, p. 38), activation within the CST increasing by 42 % as an effect of learning in the passive condition. The posterior inferior temporal region (PIT) behaved similarly, the activation within the region increasing on average by 38 %. In the case of the SF, no effects of learning were found.

Interestingly, the P2m exhibited separate areas associated with learning and intelligibility. Within the anterior superior temporal region (AST), in the passive condition, both the good-quality sentences and the second presentation of the degraded sentences elicited higher activation than the first presentation of the degraded sentences. Whereas during the first presentation of the degraded sentences, the current within the AST was on average 7.7 pA/m, during the subsequent good-quality and degraded sentences, the elicited mean currents were 10.5 pA/m and 10.6 pA/m, respectively. These results imply effects of intelligibility, that is, whether the subject was able to understand the sentences or not. The region is conspicuously similar to that identified by Scott et al. (2000) and Narain et al. (2003). Whereas Scott et al. (2000) and Narain et al. (2003) observed that the posterior parts of the superior temporal sulcus (STS) and superior temporal gyrus (STG) were more strongly activated in the presence of phonetic information, here the activation within the PIT and the CST was shown to be higher during the second presentation of the degraded sentences than during the first presentation of the degraded sentences. Similarly, whereas Scott et al. (2000) and Narain et al. (2003) associated the anterior parts of the STS with speech comprehension, here the AST was found to be more strongly activated during both the good-quality sentences and the second presentation of degraded sentences (i.e. the intelligible sentences) than during the first presentation of the degraded sentences. Similarly to the findings of Scott et al. (2000) and Narain et al. (2003), these effects were exclusive to the passive listening condition, possibly indicating that this kind of adaptation is involuntary.

Scott et al. (2000) and Narain et al. (2003) observed the effects only in the left hemisphere. As a novel finding, the effects in the present study were bilateral. Additionally, due to limitations of fMRI and PET, the previous studies were not able to describe the temporal dynamics of this process. In the present study, however, these effects were shown to be caused by cortical events occurring in the P2m time range.

Thus, whereas the anterior superior temporal regions seems to be specialized in speech comprehension, *a priori* information on the prospective degraded sentences is reflected in the elicited activation within a number of regions, extending throughout the cortex. These regions include, for the P1m, the inferior parietal regions, for the N1m, the central superior parietal region, and for the P2m, the anterior and central superior temporal regions and the posterior inferior temporal region. The effects seem to be mostly exclusive to the passive listening condition.

## 5.3 Effects of attention

Attention was found to have an influence on all of the studied responses, the P1m, N1m, P2m and SF being each affected. In the case of the transient responses and the late SF, the elicited activation was higher in the passive condition than in the active condition. The early SF, elicited in the passive condition, was weaker than that elicited in the active condition. These effects were localized to the central and posterior superior parietal regions and to the central inferior temporal region.

The effects of attention on the P1m were revealed by ROI analysis. During the P1m response, the CIT and the right-hemispheric posterior superior parietal region (PSP) showed higher activation in the passive listening condition than in the active listening condition. In the passive condition the activations within the CIT and the right-hemispheric PSP were on average 24 % and 30 % higher than those in the active condition. The effects of attention on the N1m were observed as a shift in the ECD locations. During the passive listening condition, the ECDs were more inferior, possibly indicating an increase of activity in the same inferior temporal region affected during the P1m. This could not be confirmed with the ROI analysis, however. During the P2m, the right-hemispheric ECDs showed a shift to medial direction. According to ROI analysis however, the only region affected by attention was the posterior superior parietal region (PSP). In the passive condition, the mean current within the PSP was 22 % higher than in the active condition.

The SF was also sensitive to attention. In contrast to the other responses, the early SF exhibited on average 25 % higher amplitude in the active condition than in the passive condition. This effect was present in the single-sensor measurements (i.e. in the vicinity of the auditory cortex) and is in line with the previous studies (Picton et al. 1978a, Okamoto et al. 2010). Both Picton et al. (1978a) and Okamoto et al. (2010) observed the SF to be higher in the active condition than in the passive condition. During the late SF, on the other hand, the CSP was found to be more strongly activated in the passive condition than in the active condition. In the passive condition, activation within the CSP was on average 25 % higher than in the active condition.

Both the activation of parietal areas (Tiitinen et al. 2006, Salmi et al. 2007) and auditory areas (Woldorff et al. 1993, Jäncke et al. 1999) have previously been associated with auditory attention. Here, the auditory areas were affected by attention only in the early stage (P1m), whereas attentional modulation was present in the superior parietal areas throughout the presentation of the sound. However, in contrast to the current study, in the previous studies, attending to the auditory stimuli increased the activation.

It should be noted that the passive listening condition was carried out after the active listening condition. Therefore, during the first session of the passive listening condition, the subjects already had a foresight of the prospective sentences. De-

spite this, the effects of learning and intelligibility were present during the passive condition. In future studies however, this shortcoming should be addressed e.g. by varying the order of these two conditions. The present study confirmed (in the active condition) that after listening to the good-quality sentences, the degraded sentences are better understood. Future studies could take advantage of this knowledge. Presenting the same auditory stimuli in passive condition only would potentially result in greater number of recorded epochs and a more simple experimental setting.

Also, the discrepancy between the results obtained with ECD modeling and ROI analyses is noteworthy. Whereas ROI analysis is based on the MNE data and is therefore capable of providing information on a number of sources, ECD modeling relies on the assumption on a fixed number of sources. Thus, in the present study, as the ECD model was perceived as an oversimplification, more importance was put on the results of ROI analysis. Additionally, the regions used were reasonably large. Therefore, activity in finer structures such as the STS and the STG could not be distinguished. To do so, combining the MEG information with anatomical magnetic resonance (MR) images would be useful. In addition to providing accurate spatial information, enabling to pinpoint the subject-specific locations of the regions, the fusion with MR data would also make it possible to use more accurate boundary-element models. Subsequently, the information could be used to restrict the orientations of the sources (*orientation constraints*), resulting in greater localization accuracy (Hämäläinen 2007).

# References


Aho, J. (2010), The effects of visual attention of spatial processing in human auditory cortex studied with magnetoencephalography, Master's thesis, Aalto University School of Science and Technology.

Alho, J. (2010), The role of the speech motor system in speech perception, Master's thesis, Aalto University School of Science and Technology.

Alku, P., Tiitinen, H. & Näätänen, R. (1999), 'A method for generating natural-sounding speech stimuli for cognitive brain research', *Clinical Neurophysiology* **110**(8), 1329–1333.

Anderson, J., Gilmore, R., Roper, S., Crosson, B., Bauer, M., Nadeau, S., Beversdorf, D., Cibula, J., Rogish, M., Kortencamp, S., Hugher, J., Gonzales Rothi, L. & Heilman, K. (1999), 'Conduction aphasia and the arcuate fasciculus: A reexamination of the wernicke-geschwind model', *Brain and Language* **70**(1), 1–12.

Axer, H., Gräfin, A., Berks, G. & Graf, D. (2001), 'Supra- and infrasylvian conduction aphasia', *Brain and Language* **76**(3), 317–331.

Belin, P., Zatorre, R. & Ahad, P. (2002), 'Human temporal-lobe response to vocal sounds', *Cognitive Brain Research* **13**(1), 17–26.

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

Billings, C., Tremblay, K., Stecker, C. & Tolin, W. (2009), 'Human evoked cortical activity to signal-to-noise ration and absolute signal level', *Hearing Research* **254**, 12–24.

Binder, J., Frost, J., Hammeke, T., Bellgowan, P., Springer, J., Kaufman, J. & Possing, E. (2000), 'Human temporal lobe activation by speech and nonspeech sounds', *Cerebral Cortex* **10**(5), 512–528.

Brenner, D., Lipton, J., Kaufman, L. & Williamson, S. (1978), 'Somatically evoked magnetic fields of the human brain', *Science* **199**(4324), 81–83.

Brugge, J. & Reale, R. (1985), *Cerebral Cortex: Vol. 4*, Plenum Publishing Corporation, chapter Auditory Cortex, pp. 229–271.

Budinger, E. & Heil, P. (2006), *Listening to speech: an auditory perspective*, Lawrence Erlbaum Associates, chapter Anatomy of the Auditory Cortex, pp. 91–115.

Cattermole, K. (1969), *Principles of pulse code modulation*, American Elsevier.

Coles, M. & Rugg, M. (1996), *Electrophysiology of mind: event-related potentials and cognition*, Oxford University Press, chapter Event-related brain potentials: an introduction, pp. 1–23.

Dale, A., Liu, A., Fischl, B., Buckner, R., Belliveau, J., Lewine, J. & Halgren, E. (2000), 'Dynamic statistical parameter mapping: Combining fmri and meg for high-resolution imaging of cortical activity', *Neuron* **26**(1), 55–67.

Dammers, J., Mohlberg, H., Boers, F., Tass, P., Amunts, K. & Mathiak, K. (2007), 'A new toolbox for combining magnetoencephalographic source analysis and cytoarchitectonic probabilistic data for anatomical classification of dynamic brain activity', *NeuroImage* **34**(4), 1577–1587.

Davis, A. & Merzbach, U. (1975), *Early Auditory Studies: Activities in the Psychology Laboratories of American Universities*, Smithsonian Institution Press.

Davis, H. & Zerlin, S. (1966), 'Acoustic relations of the human vertex potential', *Journal of the Acoustical Society of America* **39**(1), 109–116.

Eggermont, J. & Ponton, C. (2002), 'The neurophysiology of auditory perception: From single units to evoked potential', *Audiology & Neuro-Otology* **7**(2), 71–99.

Ele (2006), *MaxFilter, User's Guide, Software version 2.0*.

Ele (2008), *Source Modelling Software, User's Guide, Software version 5.5*.

Fabiani, M., Gratton, G. & Federmeier, K. (2007), *Handbook of Psychophysiology*, Cambridge University Press, chapter Event-Related Brain Potentials: Methods, Theory, and Applications, pp. 85–119.

Geschwind, N. & Levitsky, W. (1968), 'Human brain: Left-right asymmetries in temporal speech region', *Science* **161**(3837), 186–187.

Giard, M., Fort, A., Mouchetant-Rostaing, Y. & Pernier, J. (2000), 'Neurophysiological mechanisms of auditory selective attention in humans', *Frontiers in Bioscience* pp. d84–94.

Goldstein, B. (2007), *Sensation and Perception, Seventh Edition*, Thomson Wadsworth.

Gross, J., Kujala, J., Salmelin, R. & Schnitzler, A. (2010), *MEG: An Introduction to Methods*, Oxford University Press, chapter Noninvasive Functional Tomographic Connectivity Analysis with Magnetoencephalography, pp. 216–244.

Gutschalk, A., Patterson, R., Rupp, A., Uppenkamp, S. & Scherg, M. (2002), 'Sustained magnetic fields reveal separate sites for sound level and temporal regularity in human auditory cortex', *NeuroImage* **15**(1), 207–216.

Halgren, E., Dhond, R., Christensen, N., Van Petten, C., Marinkovic, K., Lewine, J. & Dale, A. (2002), 'N400-like magnetoencephalography responses modulated by semantic context, word frequency, and lexical class in sentences', *NeuroImage* **17**(3), 1101–1116.

Halgren, E., Liu, A., Ulber, I., Klopp, J., Heit, G. & Dale, A. (2000), From synapse to sensor: How much of the eeg/meg signal is waylaid by spatiotemporal asynchronies?, *in* 'Book of Abstracts, 12th International Conference on Biomagnetism'.

Hansen, P., Kringelbach, M. & Salmelin, R. (2010), *MEG: An Introduction to Methods*, Oxford University Press.

Hari, R. (1999), *Electroencephalography. Basic Principles, Clinical Applications, and Related Fields. Fourth Edition*, Lippincott Williams & Wilkins, chapter Magnetoencephalography as a Tool of Clinical Neurophysiology, pp. 1107–1134.

Hari, R., Aittoniemi, K., Järvinen, M., Katila, T. & Varpula, T. (1980), 'Auditory evoked transient and sustained magnetic fields of the human brain. localization of the neural generators', *Experimental Brain Research* **40**(2), 237–240.

Hari, R., Pelizzone, M., Mäkelä, J., Hällström, J., Leinonen, L. & Lounasmaa, O. (1987), 'Neuromagnetic responses of the human auditory cortex to on- and offsets of noise bursts', *International Journal of Audiology* **26**(1), 31–43.

Hari, R., Salmelin, R., Mäkelä, J., Salenius, S. & Helle, M. (1997), 'Magnetoencephalographic cortical rhythms', *International Journal of Psychophysiology* **26**(1-3), 51–62.

Helmholtz, H. (1853), 'Ueber einige gesetze der vertheilung elektrischer ströme in körperlichen leitern, mit anwendung auf die thierisch-elektrischen versuche', *Annalen der Physik* **165**(7), 353–377.

Hermann, C. & Knight, R. (2001), 'Mechanisms of human attention: event-related potentials and oscillations', *Neuroscience and Biobehavioral Reviews* **25**(6), 465–476.

Hämäläinen, M. (2007), 'Progress and challenges in multimodal data fusion', *International Congress Series* **1300**, 15–18.

Hämäläinen, M. & Hari, R. (2002), *Brain Mapping: The Methods, Second Edition*, Academic Press, chapter Magnetoencephalographic Characterization of Dynamic Brain Activation: Basic Principles and Methods of Data Collection and Source Analysis, pp. 227–253.

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

Hämäläinen, M. & Ilmoniemi, R. (1994), 'Interpreting magnetic fields of the brain: minimum norm estimates', *Medical and Biological Engineering and Computing* **32**(1), 35–42.

Hämäläinen, M., Lin, F. & Mosher, J. (2010), *MEG: An Introduction to Methods*, Oxford University Press, chapter Anatomically and Functionally Constrained Minimum-norm Estimates, pp. 186–215.

Huotilainen, M. (1997), Magnetoencephalography in the study of cortical auditory processing, PhD thesis, Helsinki University of Technology.

Inui, K., Okamoto, H., Miki, K., Gunji, A. & Kakigi, R. (2006), 'Serial and parallel processing in the human auditory cortex: A magnetoencephalographic study', *Cerebral Cortex* **16**(1), 18–30.

Jensen, O. & Hesse, C. (2010), *MEG: An Introduction to Methods*, Oxford University Press, chapter Estimating Distributed Representations of Evoked Responses and Oscillatory Brain Activity, pp. 156–185.

Jäncke, L., Mirzazade, S. & Shah, N. (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.

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

Kaas, J. & Hackett, T. (2000), 'Subdivisions of auditory cortex and processing streams in primates', *Proceesings of the National Academy of Sciences of the United States of America* **97**(22), 11793–11799.

Kaas, J., Hackett, T. & Tramo, M. (1999), 'Auditory processing in primate cerebral cortex', *Current Opinion in Neurobiology* **9**(2), 164–170.

Kaukoranta, E., Hämäläinen, M., Sarvas, J. & Hari, R. (1986), 'Mixed and sensory nerve stimulations activate different cytoarchitectonic areas in the human primary somatosensory cortex si', *Experimental Brain Research* **63**(1), 60–66.

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

Köhler, W. & Wegener, J. (1955), 'Currents of the human auditory cortex', *Journal of Cellular and Comparative Physiology* **45**(Supp. S1), 25–54.

Kutas, M., Federmeier, K., Staab, J. & Kluender, R. (2007), *Handbook of Psychophysiology*, Cambridge University Press, chapter Event-Related Brain Potentials: Methods, Theory, and Applications, pp. 555–580.

Liégeois-Chauvel, C., Musolino, A., Badier, J., Marquis, P. & Chauvel, P. (1994), 'Evoked potentials recorded from the auditory cortex in man: evaluation and topography of the middle latency components', *Electroencephalography and Clinical Neurophysiology / Evoked Potentials Section* **92**(3), 204–214.

Liikkanen, L., Tiitinen, H., Alku, P., Leino, S., Yrttiaho, S. & May, P. (2007), 'The right-hemispheric auditory cortex in humans is sensitive to degraded speech sounds', *NeuroReport* **18**(6), 601–605.

Lopes da Silva, F. (2010), *MEG: An Introduction to Methods*, Oxford University Press, chapter Electrophysiological Basis of MEG Signals, pp. 1–23.

Martin, B., Sigal, A., Kurtzberg, D. & Stapells, D. (1997), 'The effects of decreased audibility produced by high-pass noise masking on cortical event-related potentials to speech sounds /ba/ and /da/', *Journal of the Acoustical Society of America* **101**(3), 1585–1599.

Martin, B. & Stapells, D. (2005), 'Effects of low-pass noise masking on auditory event-related potentials to speech', *Ear and Hearing* **26**(2), 195–213.

May, P. & Tiitinen, H. (2010), 'Mismatch negativity (mmn), the deviance-elicited auditory deflection, explained', *Psychophysiology* **47**(1), 66–122.

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

Mäkinen, V. (2002), Methods and measurements for relating meg responses to human hearing, Master's thesis, Helsinki University of Technology.

Mäkinen, V. (2006), Analysis of the structure of time-frequency information in electromagnetic brain signals, PhD thesis, Helsinki University of Technology.

Mäkinen, V., Tiitinen, H. & May, P. (2005), 'Auditory event-related responses are generated independently of ongoing brain activity', *NeuroImage* **24**(4), 961–968.

Morita, T., Fujiki, N., Nagamine, T., Hiraumi, H., Naito, Y., Shibasaki, H. & Ito, J. (2006), 'Effects of continuous masking noise on tone-evoked magnetic fields in humans', *Brain Research* **1087**(1), 151–158.

Narain, C., Scott, S., Wise, R., ROsen, S., Leff, A., Iversen, S. & Matthews, P. (2003), 'Defining a left-lateralized response specific to intelligible speech using fmri', *Cerebral Cortex* **13**(12), 1362–1368.

Näätänen, R. & Picton, T. (1987), 'The n1 wave of the human electric and magnetic response to sound: A review and an analysis of the component structure', *Psychophysiology* **24**(4), 375–425.

Nummenmaa, A., Auranen, T., Hämäläinen, M., Jääskeläinen, I., Sams, M., Vehtari, A. & Lampinen, J. (2007), 'Automatic relevance determination based hierarchical bayesian meg inversion in practice', *NeuroImage* **37**(3), 876–889.

Okamoto, H., Stracke, H., Bermudez, P. & Pantev, C. (2010, In Press.), 'Sound processing hierarchy within human auditory cortex', *Journal of Cognitive Neuroscience* pp. 1–9.

Pandya, D. (1995), 'Anatomy of the auditory cortex', *Reviews in Neurology (Paris)* **151**(8-9), 486–494.

Pantazis, D. & Leahy, R. (2010), *MEG: An Introduction to Methods*, Oxford University Press, chapter Statistical Inference in MEG Distributed Source Imaging, pp. 245–272.

Pantev, C., Bertrand, O., Eulitz, C., Verkindt, C., Hampson, S., Schuierer, G. & 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.

Pantev, C., Eulitz, C., Elbert, T. & Hoke, M. (1994), 'The auditory evoked sustained field: origin and frequency dependence', *Electroencephalography and Clinical Neurophysiology* **90**(1), 82–90.

Pantev, C., Hoke, M., Lehnertz, K., Lütkenhöner, B., Fahrendorf, G. & Stöber, U. (1990), 'Identification of sources of brain neuronal activity with high spatiotemporal resolution through combination of neuromagnetic source localization (nmsl) and magnetic resonance imaging (mri)', *Electroencephalography and Clinical Neurophysiology* **75**(3), 173–184.

Parkkonen, L. (2010), *MEG: An Introduction to Methods*, Oxford University Press, chapter Instrumentation and Data Preprocessing, pp. 24–64.

Pasqual-Marqui, R. (2002), 'Standardized low resolution brain electromagnetic tomography (sloreta): technical details.', *Methods and Findings in Experimental and Clinical Pharmacology* **24**(Suppl D), 5–12.

Penhune, V., Zatorre, R., MacDonald, J. & Evans, A. (1996), 'Interhemispheric anatomical differences in human primary auditory cortex: Probabilistic mapping and volume measurement from magnetic resonance scans', *Cerebral Cortex* **6**(5), 661–672.

Picton, T., Bentin, S., Berg, P., Donchin, E., Hillyard, S., Johnson, R., Miller, G., Ritter, W., Ruchkin, D., Rugg, M. & Taylor, M. (2000), 'Guidelines for using human event-related potentials to study cognition: Recording standards and publication criteria', *Psychophysiology* **37**(2), 127–152.

Picton, T. & Hillyard, S. (1974), 'Human auditory evoked potentials. ii: Effect of attention', *Electroencephalography and Clinical Neurophysiology* **36**, 191–199.

Picton, T., Hillyard, S., Krausz, H. & Galambos, R. (1974), 'Human auditory evoked potentials. i: Evaluation of components', *Electroencephalography and Clinical Neurophysiology* **36**, 179–190.

Picton, T., Woods, D. & Proulx, G. (1978*a*), 'Human auditory sustained potentials. i. the nature of the response', *Electroencephalography and Clinical Neurophysiology* **45**(2), 186–197.

Picton, T., Woods, D. & Proulx, G. (1978*b*), 'Human auditory sustained potentials. ii. stimulus relationships', *Electroencephalography and Clinical Neurophysiology* **45**(2), 198–210.

Raij, T., Uutela, K. & Hari, R. (2000), 'Audiovisual intgration of letters in the human brain', *Neuron* **28**(2), 617–625.

Reite, M., Teale, P., Zimmerman, J., Davis, K. & Whalen, J. (1988), 'Source location of a 50 msec latency auditory evoked field component', *Electroencephalography and Clinical Neurophysiology* **70**(6), 490–498.

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

Salmelin, R. & Kujala, J. (2006), 'Neural representation of language: activation versus long-range connectivity', *TRENDS in Cognitive Sciences* **10**(11), 519–525.

Salmi, J., Rinne, T., Degerman, A., Salonen, O. & Alho, K. (2007), 'Orienting and maintenance of spatial attention in audition and vision: multimodal and modality-specific brain activations', *Brain Structure and Function* **212**(2), 181–194.

Scheich, H., Brechmann, A., Brosch, M., Budinger, E. & Ohl, F. (2007), 'The cognitive auditory cortex: Task-specificity of stimulus representations', *Hearing Research* **229**(1-2), 213–224.

Scherg, M., Vajsar, J. & Picton, T. (1989), 'A source analysis of the late human auditory evoked potentials', *Journal of Cognitive Neuroscience* **1**(4), 336–355.

Scott, S., Blank, C., Rosen, S. & Wise, R. (2000), 'Identification of a pathway for intelligible speech in the left temporal lobe', *Brain* **123**(12), 2400–2406.

Scott, S. & Johnsrude, I. (2003), 'The neuroanatomical and functional organization of speech perception', *TRENDS in Neurosciences* **26**(2), 100–107.

Shtyrov, Y., Kujala, T., Ilmoniemi, R. & Näätänen, R. (1999), 'Noise affects speech-signal processing differently in the cerebral hemispheres', *NeuroReport* **10**(10), 2189–2192.

Taulu, S. & Hari, R. (2009), 'Removal of magnetoencephalographic artifacts with temporal signal-space separation: Demostration with single-trial auditory evoked responses', *Human Brain Mapping* **30**(5), 1524–1534.

Taulu, S. & Kajola, M. (2005), 'Presentation of electromagnetic multichannel data: The signal space separation method', *Journal of Applied Physics* **97**(124905), 1–10.

Taulu, S., Kajola, M. & Simola, J. (2004), 'Suppression of interference and artifacts by the signal space separation method', *Brain Topography* **16**(4), 269–275.

Taulu, S. & Simola, J. (2006), 'Spatiotemporal signal space separation method for rejecting nearby interference in meg measurements', *Physics in Medicine and Biology* **51**(7), 1–10.

Taulu, S., Simola, J. & Kajola, M. (2005), 'Application of the signal space separation method', *IEEE Transactions on Signal Processing* **53**(9), 3359–3372.

Tiitinen, H., Salminen, N., Palomäki, K., Mäkinen, V., Alku, P. & May, P. (2006), 'Neuromagnetic recordings reveal the temporal dynamics of auditory spatial processing in the human cortex', *Neuroscience Letters* **396**(1), 17–22.

Tuomisto, T., Hari, R., Katila, T., Poutanen, T. & Varpula, T. (1983), 'Studies of auditory evoked magnetic and electric responses: Modality specificity and modelling', *Il Nuovo Cimento D* **2**(2), 471–483.

Vouloumanos, A., Kiehl, K., Werker, J. & Liddle, P. (2001), 'Detection of sounds in the auditory stream: Event-related fmri evidence for differential activation to speech and nonspeech', *Journal of Cognitive Neuroscience* **13**(7), 994–1005.

Vrba, J. (2002), 'Magnetoencephalography: the art of finding a needle in a haystack', *Physica C: Superconductivity* **368**(1-4), 1–9.

Wagner, M., Fuchs, M. & Kastner, J. (2004), 'Evaluation of sloreta in the presence of noise and multiple sources', *Brain Topography* **16**(4), 277–280.

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

Whiting, K., Martin, B. & Stapells, D. (1998), 'The effects of broadband noise masking on cortical event-related potentials to speech sounds /ba/ and /da/', *Ear and Hearing* **19**(3), 218–231.

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

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

Woods, D. (1995), 'The components structure of the n1 wave of the human auditory evoked potential', *Electroencephalography and Clinical Neurophysiology* **Suppl. 44**, 102–109.

Yvert, B., Fischer, C., Bertrand, O. & Pernier, J. (2005), 'Localization of human supratempiral auditory areas from intracerebral auditory evoked potentials using distributed source models', *NeuroImage* **28**(1), 140–153.