

Neural mechanisms of foreign language phoneme acquisition in early adulthood

MEG study

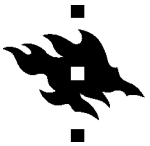
Pietari Nurmi
Master's Thesis

Cognitive Science
Faculty of Arts
University of Helsinki

Department of Neuroscience and
Biomedical Engineering
Aalto University School of Science

Instructors:
Annika Hultén (Aalto University)
Alina Leminen (University of Helsinki)

March 2020



Tiedekunta/Osasto – Fakultet/Sektion – Faculty Humanistinen tiedekunta		
Tekijä – Författare – Author Sampsu Pietari Nurmi		
Työn nimi – Arbetets titel – Title Neural mechanisms of foreign language phoneme acquisition in early adulthood – MEG study		
Oppiaine – Läroämne – Subject Kognitiotiede		
Työn laji – Arbetets art – Level Pro gradu -tutkielma	Aika – Datum – Month and year maaliskuu 2020	Sivumäärä – Sidoantal – Number of pages 52
Tiivistelmä – Referat – Abstract <p>Tämän tutkimuksen tavoitteena on selvittää omaan äidinkieleen kuulumattomien foneemikontrastien oppimisen mekanismeja nuorilla aikuisilla neurofysiologisten ja behavioraalisten menetelmien avulla. Perinteisesti kielen foneettisen avaruuden omaksumisen on ajateltu tapahtuvan ensisijaisesti varhaislapsuuden kielellisten herkkyyskausien aikana, jonka jälkeen uusien foneemien oppiminen on haastavaa. Myöhemmät tutkimukset ovat kuitenkin osoittaneet, että vieraiden foneemien omaksuminen on mahdollista myös aikuisiällä. Uusien foneemikategorioiden muodostuminen vaatii aivoissa solutason plastisia muutoksia. Aivojen kykyä erotella läheisesti toisiaan muistuttavia foneemikategorioita kielenprosessoinnin varhaisella tasolla on tutkittu neurofysiologisista menetelmistä esimerkiksi tapahtumasidonnan poikkeavuusnegatiivisuusvasteen (eng. <i>mismatch negativity</i>, MMN) avulla. MMN-vaste, tai sen magneettinen vastine MMNm, syntyy seurauksena muutoksiin sensorisessa havaintoympäristössä. Tutkimuksissa lyhyenkin auditiivisen harjoittelujakson on havaittu vahvistavan aivojen kykyä erotella läheisesti toisiaan muistuttavia vieraita foneemeja ja voimistavan MMN- ja MMNm-vasteita. Tässä tutkimuksessa vieraan kielen foneettisen oppimisen neuraalista perustaa ja oppimisen aiheuttamia plastisia muutoksia aivoissa tutkittiin magnetoenkefalografialla (MEG) neuromagneettisten tapahtumasidonnan vasteiden (erityisesti MMNm) avulla.</p> <p>Tutkimuksessa mitattiin 20 suomalaista koehenkilöä, joiden tehtävänä oli oppia erottelamaan akustisesti toisiaan läheisesti muistuttavia venäjän kielen frikatiiveja Ш /ʃ/ ja Щ /ɕ(:)/. Erottelukykyä mitattiin ensin behavioraalisella tehtävällä, jossa koehenkilöille toistettiin nauhoitettuja venäjänkielisiä epäsanaminimipareja, jossa sanan ensimmäistä foneemia varioitiin. Koehenkilöiden tehtävänä oli vastata, kuulivatko he sanoissa eroa. Samoja kuuloärsykeitä toistettiin koehenkilöille sen jälkeen passiivisessa MEG-tehtävässä, jossa testattiin aivojen kykyä havaita ero ärsykkeissä ilman, että niihin kiinnitetään huomiota (koehenkilöt katselivat samalla äänetöntä elokuvaa). Mittauksen jälkeen koehenkilöt harjoittelivat foneemien erottelua kotona noin viikon ajan tietokoneavusteisen oppimispelin avulla, jonka jälkeen heidät mitattiin uudelleen. MEG-signaalien lähdemallinnusta varten koehenkilöiden aivoista otettiin myös rakenteelliset magneettikuvat.</p> <p>Tutkittavien foneemien behavioraalinen erottelukyky oli selvästi tuttuja kontrollifoneemeita heikompa. Erottelukyky vaikutti paranevan harjoittelun seurauksena hieman, mutta ero ei ollut tilastollisesti merkitsevä. Hypoteesien vastaisesti tilastollisesti merkitseviä MMNm-vasteita ei löydetty ennen eikä jälkeen harjoittelun, eikä muissakaan auditorisissa MEG-vasteissa tai niiden neuraalisten lähdevirtojen voimakkuuksissa tai jakaumassa ollut tilastollisesti merkitsevää eroa mittauskertojen välillä. Yksilölliset erot oppimisessa olivat kuitenkin suuria. Koehenkilöillä, joilla behavioraalinen erottelukyky parani harjoittelun myötä, oli silmämääräisesti havaittavissa hypoteesien mukaista vahvistumista auditorisissa vasteissa. Vaikka efekti oli erittäin pieni eikä tilastollisesti merkitsevä, vastaavaa ei havaittu epäoppijoilla eikä kontrollitilanteessa.</p> <p>Tässä tutkimuksessa ei kyetty replikoimaan aiempien tutkimusten tuloksia foneemien omaksumisesta aikuisiällä. Vaikka on todennäköistä, että tietyt metodologiset heikkoudet (mm. vähäinen ärsykkeiden määrä MEG-tehtävässä, haastavat ärsykkeet) vaikuttivat tulosten merkitsevyyteen, voidaan tämän tutkimuksen valossa aiempien tutkimustulosten yleistettävyyttä kyseenalaistaa.</p>		
Avainsanat – Nyckelord – Keywords foneettinen oppiminen, vieras kieli, havainto-oppiminen, auditiivinen harjoittelu, auditorinen plastisiteetti, MEG, MMNm		
Säilytyspaikka – Förvaringställe – Where deposited Helsingin yliopiston kirjasto – HELDA / E-thesis		
Muita tietoja – Övriga uppgifter – Additional information		



Tiedekunta/Osasto – Fakultet/Sektion – Faculty Faculty of Arts		
Tekijä – Författare – Author Sampsa <u>Pietari</u> Nurmi		
Työn nimi – Arbetets titel – Title Neural mechanisms of foreign language phoneme acquisition in early adulthood – MEG study		
Oppiaine – Läroämne – Subject Cognitive Science		
Työn laji – Arbetets art – Level Master's thesis	Aika – Datum – Month and year March 2020	Sivumäärä – Sidoantal – Number of pages 52
Tiivistelmä – Referat – Abstract <p>The aim of this study is to examine the learning mechanisms and acquisition of non-native phoneme contrasts in young adults using neurophysiological and behavioral methods. According to the traditional view, acquiring novel phonemes after the sensitive periods in the early childhood is very difficult. However, later findings have shown that foreign phoneme contrasts can be learned at a later age, too. Acquiring new phonemic categories requires neuroplastic changes in the brain. Neurophysiological studies have examined the brain's ability to differentiate between closely related phonemic categories at the early stage of spoken language processing by measuring, for example, event-related mismatch negativity responses (MMN). MMN, or its magnetic equivalent MMNm, is elicited when the brain registers a difference in a repetitive sensory stimulus. Studies have shown that even a moderate amount of auditory training with closely related foreign phonemes improves the brain's ability to discriminate between them resulting in enhanced MMN or MMNm responses. In this experiment the neural mechanisms of foreign language phoneme acquisition and the learning-related neuroplastic changes were studied using magnetoencephalography (MEG) and neuromagnetic evoked responses (MMNm in particular).</p> <p>20 Finnish subjects were measured in the experiment. Their task was to learn to differentiate between acoustically closely related Russian fricatives Ш /ʃ/ and Ш /ɕ(:)/. The subjects' differentiation skills were first tested in a behavioral task where Russian pseudoword minimal pairs were presented to them auditorily. The first phoneme in the word pairs was varied and the subjects had to report whether they heard a difference between the words or not. The same stimuli were then presented in a passive MEG task where the brain's change detection responses were tested in an unattended situation as the subjects were watching a silent film. After the measurement the subjects practiced the phonemes at home for approximately one week by playing a learning game by computer. After training they were measured again. Structural magnetic resonance images of the subjects' brain were also measured for MEG source localization purposes.</p> <p>Behavioral discrimination ability of the experimental phonemes was considerably worse than with familiar control phonemes. The discrimination skills seemed to improve by training, but the difference was not statistically significant. Contrary to the hypotheses, statistically significant MMNm responses were not found before or after training. No significant differences were found in other auditory MEG responses or their neural source current distributions between the measurements either. However, individual differences in learning were sizeable. For the subjects who improved their performance in the behavioral task a modest training-related boost in the auditory responses supporting the hypotheses could be observed. Although very small and statistically insignificant, the effect was opposite for control stimuli and did not exist in the non-learner group suggesting some sort of change in neural processing in the learner group.</p> <p>This study was not able to replicate the findings from various previous studies on phoneme acquisition in adulthood. Although it is likely that certain methodological limitations (e.g. small number of stimulus repetitions, challenging stimuli) affected the significance of the results, based on this study the generalizability of some of the previous findings can be called into question.</p>		
Avainsanat – Nyckelord – Keywords phoneme acquisition, foreign language, perceptual learning, auditory training, auditory plasticity, MEG, MMNm		
Säilytyspaikka – Förvaringställe – Where deposited Helsinki University Library – HELDA / E-thesis		
Muita tietoja – Övriga uppgifter – Additional information		

PREFACE	1
1. INTRODUCTION	2
2. BACKGROUND	4
2.1. Neuroimaging methods & language.....	4
2.1.1. MEG and other electrophysiological methods	4
2.1.2. MEG signals and source localization.....	5
2.1.3. Event-related evoked responses.....	6
2.2. Neural mechanisms of phoneme perception.....	7
2.2.1. Phonemic prototypes	7
2.2.2. Neural representations of phonemes	7
2.2.3. Neural correlates of phoneme perception	8
2.3. Foreign language acquisition.....	8
2.4. Foreign phoneme learning and brain plasticity	9
2.4.1. Training-related neuroplastic changes.....	10
2.5. Motivation for current study.....	11
2.5.1. Stimulus phonemes	11
2.5.2. Studying phonemic processing with MEG	13
2.6. Research questions and hypotheses.....	13
3. METHODS	15
3.1. Subjects	15
3.2. Experimental procedure	15
3.3. Stimuli	16
3.3.1. Editing	17
3.3.2. Training stimuli	19
3.4. Measurement sessions	19
3.4.1. Behavioral discrimination task.....	20
3.4.2. MEG oddball task.....	21
3.4.3. MRI session	22
3.5. Gamified training period	22
3.6. MEG data analysis	23
3.6.1. Data & preprocessing	23
3.6.2. Event-related fields.....	24
3.6.3. Source modelling.....	25
3.7. Statistical analyses.....	26
3.7.1. Behavioral data analysis	26
3.7.2. MEG statistical analysis.....	26
4. RESULTS	28

4.1. Behavioral learning results	28
4.2. MEG sensory ERFs	29
4.3. MEG source modelling	31
4.3.1. <i>MMNm responses</i>	32
4.3.2. <i>Training-related effects</i>	34
4.4. Interactions between MEG and behavioral results	34
5. DISCUSSION	38
5.1. Behavioral learning results	38
5.2. MEG results.....	39
5.2.1. <i>MMNm responses</i>	39
5.2.2. <i>Relationship between MEG results and learning</i>	40
5.2.3. <i>Late activation</i>	41
5.3. Methodological limitations.....	42
5.3.1. <i>Stimuli</i>	42
5.3.2. <i>Number of repetitions</i>	44
5.3.3. <i>Time windows of interest</i>	45
5.4. Conclusions and future directions	45
6. REFERENCES.....	48

Preface

The study reported in this thesis was conducted in the Imaging Language research group at the Department of Neuroscience and Biomedical Engineering, Aalto University School of Science as part of a research project *Individual differences in novel phoneme learning ability*. The project was funded by the Academy of Finland.

The background work for the research was done by PhD Annika Hultén (principal investigator). I joined the project right from the beginning as a research assistant and a humble 2nd year student (at this time I had no idea that the very project would one day give birth to this thesis). Being part of the project for almost the entirety of its course, I have had the opportunity of being involved in (or even responsible for) the majority of the practical work from designing the experiments, recording the stimuli, programming the stimulus sequences and piloting to the actual measurements and analysis.

Over the course of this work I received help from various people. I am extremely grateful for my instructor Annika Hultén for the opportunity of being a part of this project, for being my mentor and above all, the best guide to the world of MEG. The years we put into this project have taught me more than any course work ever could have. I want also to thank MD, PhD Hanna Renvall for providing valuable insight and assistance during the course of the study, Clinical Neurophysiology Nurse Mia Illman for helping in the MEG measurements, AMI Centre staff for the hours we spent together in the MRI lab and all the fellow NBE research assistants and of course Prof. Riitta Salmelin and the rest of the Imaging Language group for providing the best and most supporting working environment one could hope for. I also want to express my gratitude to the Cognitive Science staff at the University of Helsinki and most of all to my thesis instructor PhD, Docent Alina Leminen for sacrificing her time and being there when I almost lost the hope. Without you this thesis would never have been finished.

Pietari Nurmi

March 2020

1. Introduction

Phoneme is the smallest unit of spoken language affecting the semantical properties of a word. Each language has a set of approximately 40 inherent phonemes that differentiate words and syllables from one another. This language specific phonemic inventory is acquired in early childhood during the *sensitive period* of phonetic learning which often occurs before the first birthday (Kuhl, 2010). The neural representations of the *phonemic categories* of child's mother tongue develop already at the age of 6–12 months (Cheour et al., 1998). During this period children learn to categorize the speech sounds they hear into the phonemic categories present in their mother tongue. These categories shape the way we hear speech sounds. We have no troubles of hearing even the slight acoustic distinctions between the phonemes of our mother tongue, but foreign phonemes can prove a challenge. For example, English phonemes /r/ and /l/ are difficult for Japanese speakers (Miyawaki et al., 1975) and Hindi contrast between /t/ and /ʈ/ is hard to perceive for English speakers (Werker & Tees, 1984).

Most of the languages in the world have at least some sort of a sibilant or 's-sound' in their phonemic inventory (Maddieson, 1984). Children exposed to English language quickly learn to differentiate between the words *sea* and *tea* based on a change in single phoneme. Many languages, English included, have several different variants of sibilant sounds in their phonemic inventory. Slavic languages are particularly interesting in this sense. Russian and other Slavic languages are famous for the number of different sibilant sounds some of which closely resemble each other by their acoustic properties (Zygis, 2003). Russian children are exposed to a variety of different sibilant sounds in their everyday perceptual environment. From early on they learn to differentiate between sharp sounding *voiceless alveolar sibilant fricative* /s/ (Cyrillic letter *с*, similar to the pronunciation of ⟨s⟩ in *sea*), softer *voiceless retroflex sibilant fricative* /ʂ/ (Cyrillic letter *ш*, resembles the pronunciation of ⟨sh⟩ in *sheep*), hissing *voiceless alveolo-palatal sibilant fricative* /ɕ(:)/ (Cyrillic letter *щ*, pronounced as a mixture of ⟨sh⟩ and ⟨ch⟩) and other perceptually more or less similar speech sounds present in the phonemic inventory of Russian language. In contrast, Finnish language has only one sibilant phoneme inherently: the Finnish 's' /s/ is an equivalent of the Russian 'с'. After the sensitive period of language acquisition the child's ability to perceive differences between phonemes that do not belong in the phonemic inventory of his/her mother tongue deteriorates (Kuhl, 2010). As Finnish children are exposed to a rather limited variety of different sibilants in their early

childhood, their ability to differentiate between, for example, Russian sibilants without training remains poor, even later in life.

When trying to learn a foreign spoken language, unfamiliar phonemes can prove quite a challenge. Overlooking meaningful nuances in speech can sometimes result in serious misunderstandings. Particularly challenging are so called **minimal pairs** where two words with different meanings are separated from each other by a single phoneme. For example, Russian words *шёлк* [ʂɔlk] (*silk*) and *щёлк* [ɕ:ɵlk] (*click*) differ from each other only by a minor change in the first fricative. Even though young children have far superior ability to learn languages and acquire phonemic categories, evidence from neurophysiological and behavioral studies suggests that with proper training novel phoneme categories can be learned also at a later age (e.g. Kujala & Näätänen, 2010; Winkler et al., 1999). In this thesis and experiment we were interested in the neural mechanisms underlying foreign phoneme acquisition. We chose to investigate the perceptual acquisition of Russian sibilant phoneme contrast ш–щ in young Finnish adults using minimal pairs in a gamified training paradigm. Training-related effects and neuroplastic changes were studied using magnetoencephalography (MEG) and behavioral measurements.

2. Background

2.1. Neuroimaging methods & language

Noninvasive neuroimaging methods have been developing rapidly during the last few decades. This has made it possible to study the neural mechanisms underlying early language processing and phoneme perception extensively in both adults and infants. Temporally accurate electrophysiological methods such as electroencephalography (EEG) and magnetoencephalography (MEG) are well suited for studies on language processing (Kuhl, 2010). These methods measure the electrical activity of the brain seen outside the scalp. EEG measures the sum of electrical currents produced by neuronal activity whereas MEG measures the magnetic fields produced by these currents.

The following two chapters are essentially based on Hari & Puce's *MEG-EEG Primer* (2017), unless mentioned otherwise.

2.1.1. MEG and other electrophysiological methods

Electrical currents induced by neuronal activity in the brain produce both electrical potentials and magnetic fields that can be measured non-invasively outside the head. Probably the most traditional and commonly used method of measuring brain's electrical activity is electroencephalography, or EEG, which measures electrical potentials seen on the scalp. Magnetoencephalography, or MEG, is in many ways similar to EEG but instead of measuring electrical potentials it measures the magnetic fields produced by the electrical currents in the brain. Both of these electrophysiological measurement techniques capture mostly postsynaptic activity of cortical pyramidal neurons. Pyramidal neurons, unlike nonpyramidal, are aligned and perpendicularly oriented towards the cortex allowing currents to sum into measurable macroscopic net currents. One of the main advantages of electrophysiological methods is that they measure neuronal activity directly in contrast to indirect methods such as functional Magnetic Resonance Imaging (fMRI) which reflects neuronal activity based on relatively slow metabolic changes in the brain. Both MEG and EEG are temporally precise and can record and differentiate brain activity variations at a sub-millisecond accuracy. Therefore, they are most useful when studying time sensitive event-related sensory responses or dynamic neurophysiological rhythms in the brain. Language processing, especially at the lower levels,

is rapid, hence high temporal accuracy is required in order to study different stages of processing.

Both MEG and EEG have their advantages and disadvantages. The main difference between them originates from the recording technique. It is known that radial source currents inside a spherical conductor produce no magnetic field visible outside the sphere (Hämäläinen, Hari, Ilmoniemi, Knuutila, & Lounasmaa, 1993). Due to the fact that human head is an almost spherical conductor, radial source currents in the brain are not visible in MEG signal. Therefore, MEG is selective towards tangential source currents and tangential components of tilted currents. All currents at the center of a sphere are radial, which is one of the reasons why MEG is not particularly sensitive to deep currents either. EEG detects both radial and deep currents also but because it measures electric potentials from the scalp the signal gets distorted by differing electrical conductivities of the skin, skull and cerebrospinal fluid. This spreading or “*smearing*” of the potentials makes EEG signal source estimation difficult. However, different conductivities do not affect the magnetic fields produced by the same source currents making MEG signal much less distorted than EEG. The main advantage of this is MEG’s higher spatial resolution.

2.1.2. MEG signals and source localization

Magnetic fields produced by the brain are weak (measured in femto- or picoteslas) which is why extremely sensitive superconductive sensors operating near the absolute zero temperature are needed. SQUID (superconducting quantum interference device) is the most commonly used sensor type. The sensors form flux transformers, or pickup coils, which are used to measure the magnetic fields. *Gradiometers* measure the difference of magnetic flux using two or more measurement points. They are sensitive to signals from nearby sources and resilient to distant sources which improves their signal-to-noise-ratio but also reduces their ability to detect deep signal sources from the brain. *Magnetometers* are just single sensors without any compensation coil. They are sensitive to distant and deep sources but at the same time more prone to noise and various signal artifacts.

The ability to accurately estimate the location and strength of signal source currents in the brain is one of the main advantages of MEG over EEG. Source estimates can be computed when the magnetic field is measured from multiple locations outside the head. However, computational source estimation is not totally unambiguous due to so called *inverse problem* (Helmholtz,

1853). Because of magnetically silent currents (currents that are undetectable outside the head, radial currents in MEG) the exact same magnetic field detected by MEG can actually be produced by multiple different source current distributions. However, reliable MEG signal source estimations can be computed by analyzing the physiological plausibility of different source candidates based on previous knowledge of brain anatomy and existing evidence from other neuroimaging methods. In order to reach the desired spatial resolution MEG data needs to be co-registered with individual structural Magnetic Resonance (MR) images of subject's brain (Singh, Holliday, Furlong, & Harding, 1997). Similar source estimation method can be applied in EEG also, but it requires additional presumptions and is less reliable and accurate due to signal spreading.

2.1.3. Event-related evoked responses

As phonemes and speech sounds are time-locked and can be easily controlled they are often studied using event-related evoked responses of electrophysiological brain imaging data. In EEG these responses are called event-related potentials (ERPs), and the MEG equivalent are event-related fields (ERFs) (Luck, 2014). First EEG or MEG signal is recorded during a (linguistic) task. Then the data are divided into stimulus event-related segments. Signal-to-noise ratio (SNR) of single data events is quite poor which is why in a traditional ERP/ERF experiment similar stimulus events (e.g. certain phonemes) are repeated multiple times during the task. Finally, all data segments related to a specific stimulus events are averaged into event-related evoked responses. In order to draw reliable conclusions between brain activity and certain stimulus parameters, ERP/ERF experiments need to be strictly controlled (Luck, 2014).

One of the most popular evoked responses paradigms in research on early language processing is mismatch negativity (MMN). MMN is a well-known electrophysiological component that is associated with brain's automatic change detection system (Näätänen, Paavilainen, Rinne, & Alho, 2007). It is usually studied with an oddball task where a sequence of repetitive standard stimulus is followed by a differing deviant stimulus. MMN response is elicited if the deviant stimulus is found incongruent with the memory representation of the preceding stimulus. It usually peaks 150–250ms after deviant stimulus onset. Greater deviation between standard and deviant stimuli results in larger MMN amplitude and earlier peak latency (Näätänen et al., 2007). Due to its nature MMN can be used to study the neural basis and plasticity of stimulus discrimination (Kujala & Näätänen, 2010). MMN can be measured both with EEG and MEG. MEG equivalent for MMN response is marked either MMNm or MMF (mismatch field).

2.2. Neural mechanisms of phoneme perception

2.2.1. Phonemic prototypes

Kuhl (1991) suggested that each phonemic category of a person's mother tongue is formed around a 'phonemic prototype' stored in long term memory. This prototype represents how that particular phoneme should sound in an ideal condition. According to the prototype theory the distinctions between phonemic categories are therefore based on a representation of the characteristics and attributes of an ideal phoneme rather than on their boundaries against other categories. This approach has been supported by the evidence of a 'perceptual magnet' effect. Kuhl (1991) and subsequent studies showed that unrepresentative phonemes (those acoustically further away from the prototype) were perceived closer to the prototype phoneme than they actually were by their physical characteristics.

The neural basis of perceptual magnet effect has been studied with EEG. Aaltonen, Eerola, Hellström, Uusipaikka, & Lang (1997) found that MMN responses for synthetic phoneme pairs close to the prototype phoneme were smaller than for other non-prototypical phoneme pairs while the magnitude of the physical change was the same in both pairs. This indicates that the pairs affected by the proximity of the prototype phoneme were perceived closer to each other which was also supported by behavioral discrimination accuracy results. When trying to learn foreign languages the perceptual magnet effect can lead to misunderstandings if two different, but acoustically closely related foreign phonemes are incorrectly perceived as a member of the same category. This is often the case if two or more foreign phonemes (e.g. Russian *u* and *υ*) closely resemble a single prototype phoneme of one's native language (e.g. Finnish *s*).

2.2.2. Neural representations of phonemes

Phonemic prototypes of one's mother tongue are developed in early childhood (Cheour et al., 1998; Kuhl, 2010; Kuhl, Williams, Lacerda, Stevens, & Lindblom, 1992). Sensory experience gained by long-term exposure to spoken language establishes memory traces of auditory patterns unique to that language in the brain. Regularly heard speech components are stored in the long-term memory where they will eventually form the prototypic memory traces representing phonemic categories. When processing speech stimuli these traces are activated and used for discriminating, analyzing and categorizing even the smallest significant speech components into phonemes. These traces are language-dependent and do not activate when

processing equally complex non-speech auditory stimuli (Cheour et al., 1998; Näätänen et al., 1997).

Evidence of these kind of memory traces in the brain has been provided by EEG and MEG measurements alike (Näätänen et al., 2007). For example, Näätänen et al. (1997) studied Estonian and Finnish phonemes /*ö*/, /*o*/ and /*õ*/ with Finnish and Estonian subjects using EEG. Phonemes /*ö*/ and /*o*/ appear in both languages whereas /*õ*/, which stands acoustically between these two, is unique for Estonian. These phonemes were then presented as deviant stimuli in an oddball task where the standard was always /*e*/ (which also appears on both languages). Estonians showed greater MMN amplitudes as the physical difference between the deviant and standard phoneme grew larger. Finns performed similarly on /*ö*/ and /*o*/ but MMN amplitude for /*õ*/ was decreased, as it had no representation in their long-term memory. Similar results have found, for example, with French and Japanese (Dehaene-Lambertz, Dupoux, & Gout, 2000) and English and Japanese (Phillips et al., 1995) subjects.

2.2.3. Neural correlates of phoneme perception

In right-handed population language functions in the brain are typically lateralized to the left hemisphere (Knecht et al., 2000). This applies to phonemic processing also. For example, Tervaniemi et al., (2000) investigated perception of phonemes and musical chords with Positron Emission Tomography (PET). They found that phoneme stimuli activated left superior temporal and middle temporal gyri in contrast to musical stimuli which activated right superior temporal gyrus. In another study Imada et al., (2006) studied infants' phoneme detection with MEG and MMNm responses. For 12-month infants two MMNm generators were found: one in superior temporal region and another in inferior frontal region. Multiple studies link superior temporal gyrus (STG) in phoneme perception. STG is associated with phoneme processing and phonetic feature encoding (Buchsbaum, Hickok, & Humphries, 2001; Mesgarani, Cheung, Johnson, & Chang, 2014). There is even evidence of neural representations of phonemic categories in the posterior STG (Chang et al., 2010).

2.3. Foreign language acquisition

According to the sensitive period hypothesis there are specific time windows during which children or infants' ability to acquire language (or different aspects of it, such as phonemes) is

heavily emphasized (Kuhl, 2010). If a child receives no proper language input during the first years, his/her language skills will never develop to their full potential. Popular example of this are the documented cases of feral and deaf children who have not been exposed to language in their early childhood and remain linguistically deprived even after they receive training later in their childhood (Vyshedskiy, Mahapatra, & Dunn, 2017).

While sensitive period hypothesis is widely accepted when considering first language acquisition, many researchers have questioned it in the context of second language acquisition. There is a strong relationship between second language proficiency and the age of acquisition: the earlier the learning happens, the more fluent the language skills tend to be (Kuhl, 2010). However, this deterioration of language skills seems to be progressive and does not follow any maturational milestones (Birdsong, 2005). Furthermore, multiple studies in multiple languages suggest that native-like performance in second language can be achieved even if the training begins in the late teens or early adulthood (Birdsong, 2005; Singleton, 2001). ‘Native-likeness’ in these studies is usually measured based on grammatical proficiency and pronunciation. It is noteworthy that the individual differences between ‘late learners’ are considerable and usually just 5–15% of the sample can reach the native-like performance, depending on the study (Birdsong, 2005). Nonetheless, this seems to indicate that second language attainment after the early childhood, even at a native-like level, is far from impossible, unlike the sensitive period hypothesis (let alone its more profound version *critical* period hypothesis) would suggest. This raises a question whether some kind of plastic changes take place in the brain along with learning, especially in early language processing.

2.4. Foreign phoneme learning and brain plasticity

As discussed earlier, early childhood exposure to spoken language eventually forms memory representations of phonemic categories and stores them in the long-term memory. In a particularly interesting study Winkler et al. (1999) investigated Hungarian immigrants who had moved to Finland as adults or in their youth but spent their childhood years in Hungary. In Finnish language sounds /e/ and /æ/ form separate phonemic categories *e* and *ä*, unlike in Hungarian where they belong in the same category (in Budapest dialect). Winkler et al. found out that those Hungarians (from Budapest area) who had learned to speak fluent Finnish after the immigration were able to discriminate between these phonemes in a behavioral task.

Furthermore, their discrimination ability was seen in the brain also through strong MMN responses similar to those recorded with the Finnish control group. Winkler et al. concluded that these Hungarians had acquired cortical memory representations for novel Finnish phonemes. Those Hungarians who had not learned to speak Finnish could not discriminate between the phonemes behaviorally and no significant MMN responses were found.

2.4.1. Training-related neuroplastic changes

Evidence of similar learning-related neuroplastic changes in the brain have been found also in more controlled training study paradigms. For example, Kraus, Carrell, King, Tremblay, & Nkol (1995) trained adult subjects to differentiate between two slightly differing artificially created /da/ syllables. They found that as subjects' behavioral discrimination accuracy got better also their MMN amplitudes grew larger. In a similar study Tremblay, Kraus, Carrell, & McGee (1997) varied voice-onset times (VOT) in synthetic /ba/ syllables. These VOT cues are not relevant in English language and were therefore unfamiliar for the English subjects. Before training subjects could not differentiate between different /ba/ sounds but after training their discrimination accuracy improved and larger MMN responses were recorded, consistent with Kraus' et al. results. Control group who did not receive any training performed similarly in both measurements and no significant changes in MMN responses were found. Interestingly, in the experimental group the training effect acquired with /ba/ stimuli was generalized to spectrally different /da/ syllables with similar VOT cues even though these sounds were not used in the training. The effect was found both in behavioral discrimination and MMN responses. These results demonstrate that the auditory language processing system is responsive to training even in adults and that the training effects are generalizable to different stimuli.

Similar results have been acquired with natural language stimuli also. For example Menning, Imaizumi, Zwitserlood, & Pantev (2002) taught young German adults to differentiate between Japanese words whose duration was altered according to the Japanese 'mora' word segmentation system. These duration differences were therefore foreign and unfamiliar for the Germans. The subjects were measured with MEG and did a behavioral task before and after training. The duration discrimination accuracy and reaction times of the German subjects improved after training and stronger MMNm was recorded. According to Menning et al. (2002) the results suggest a strong relationship between perceptual learning and cortical plasticity. In another study Ylinen et al. (2009) investigated Finnish and English subjects with EEG. In Finnish language phonemic duration cues are highly relevant and therefore Finns tend to rely

on them heavily also when processing foreign speech sounds. When Ylinen et al. compared the use of spectral and duration cues in English vowels /i/ and /I/ (e.g. beat and bit) the Finnish subjects relied more on the vowel duration than the English subjects who used the spectral difference as a cue. However, after Ylinen et al. trained the Finnish subjects with a minimal pair identification task they were able to use the spectral cues more accurately. Also, their MMN responses for spectral cues grew in amplitude suggesting plastic changes in pre-attentive phonetic cue weighting at the early stages of language processing in the brain.

It also seems that the physiological changes in the brain, indicated by MMN responses, take place quite fast with moderate amount of training. These changes happen before any behavioral learning effect can be seen. Tremblay, Kraus, & McGee (1998) showed (using the same stimuli as Tremblay et al., 1997) that MMN amplitudes were significantly increased after only one training session with all subjects. However, there were significant individual differences in the amount of additional training required before the subjects showed signs of perceptual learning. Tremblay et al. concluded that there are considerable individual differences in how well people can use pre-attentive physiological cues in the learning process.

2.5. Motivation for current study

The aim of this thesis and study was to address general neural mechanisms of foreign phoneme acquisition in adulthood. In terms of controllability, artificially created stimuli would have been a sound option. Instead, we wanted to improve the ecological validity of the study and decided to use spoken phonemes of a natural language. We also thought that from a subject's perspective it would be more sensible and motivating to study a real language than artificial phoneme contrasts. We chose to use Russian language and Finnish subjects mostly out of convenience. It was relatively easy to find native Russian speakers in Finland and we had a Russian linguistic contact who was willing to help us with the stimuli. Furthermore, Russian phonology differs considerably from Finnish and most Finns are rarely exposed to Russian speech in their everyday sound environment.

2.5.1. Stimulus phonemes

Sibilants in Slavic languages can be categorized in two to four *sibilant categories* based on the place of articulation. In Russian sibilants can be categorized at least in dental, retroflex and

alveolo-palatal (or postalveolar) sibilants. (Padgett & Zygis, 2007). Notice the difference between a *sibilant category* (place of articulation) and a *phonemic category* mentioned earlier. Several different phonemes (phonemic categories) can exist within a single sibilant category. Consonants in general can be categorized into fricatives and affricates based on how the articulators are placed. Fricatives are produced by passing air through a narrow channel while affricates begin with a full stop. Both of these can originate from the same place of articulation but still act as a distinctive feature between two different phonemic categories.

In this study and thesis, we decided to concentrate on two Russian sibilant phonemes, represented with the Cyrillic letters ш and щ, which closely resemble each other by their physical characteristics (Figure 1). Ш /ʃ/ (voiceless retroflex sibilant fricative) is represented with ⟨ʃ⟩ in the International Phonetic Alphabet (IPA) but is pronounced somewhat similar to the ⟨sh⟩ in the English word *sheep* and therefore often romanized as ⟨sh⟩. Щ /ɕ(:)/ (voiceless alveolo-palatal sibilant fricative) is represented with ⟨ɕ⟩ in the IPA and romanized as ⟨shch⟩. It has no direct English equivalent, but it resembles a sound made by combining sounds ⟨sh⟩ (as in *sheep*) and ⟨ch⟩ (as in *cheek*). In Finnish language there has traditionally been only one phoneme category /s/ (represented with the letter *s*) for these types of sibilant sounds. Phoneme /ʃ/ (represented either with the letter combination ‘*sh*’ or letter *š*) that resembles Russian ш, exists in certain loanwords, such as ‘*shakki/šakki*’ (chess), but it is the least common phoneme in Finnish language and its status in Finnish phonology is unestablished (Hakulinen et al., 2004). We hypothesized that untrained adult Finns categorize these two Russian phonemes into the same phonemic category making it hard to discriminate between them. According to our pilot study (12 Finnish and 5 Russian subjects) differentiating between them tends to be extremely difficult for untrained Finns especially in the context of words but also when comparing isolated phonemes. However, we expect that practicing phoneme discrimination for several days with a computer game does affect the discrimination accuracy and neuromagnetic evoked responses measured after training. To our knowledge Russian ш and щ phoneme discrimination has not been previously studied on Finnish (or any other) subjects.

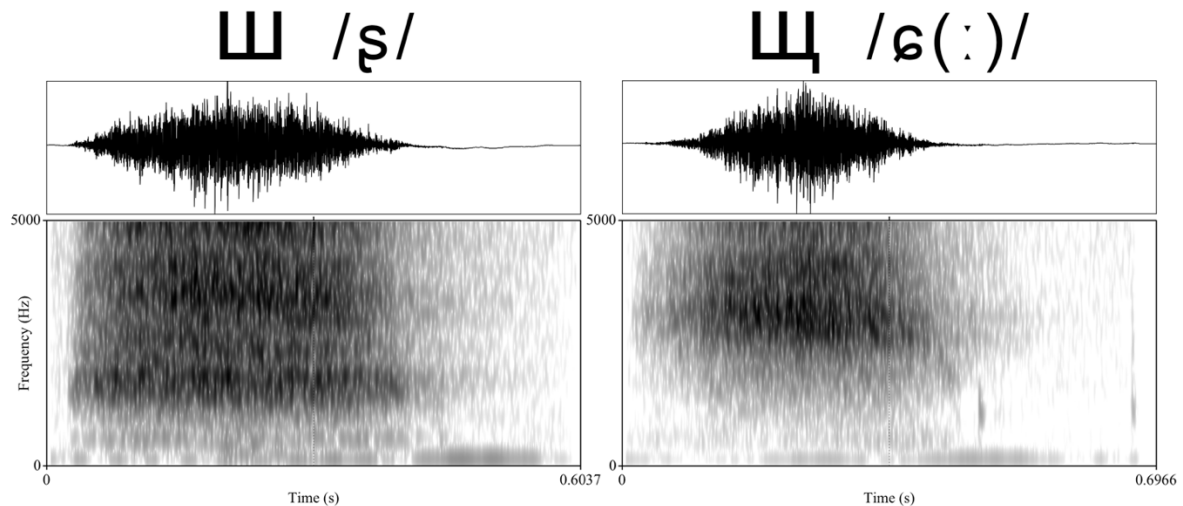


Figure 1. Spectrograms and waveforms of separately pronounced *u* and *u:* phonemes. The two phonemes resemble each other closely by their spectral properties. In the context of words *u:* is usually pronounced a little longer than *u*.

2.5.2. Studying phonemic processing with MEG

In EEG separating the left and right hemisphere sources is difficult because of the spreading of the signal. With MEG different sources can be identified easily, which makes MEG especially handy when studying brain functions with more than one potential signal source. When studying heavily lateralized brain functions, such as language, potential signal sources can be restricted to one hemisphere only. This results in improved SNR and more reliable and accurate interpretations of the brain functionality. Also, the temporal MMN sources, associated with phoneme detection, are tangentially oriented which is ideal for MEG gradiometers (Kujala, Tervaniemi, & Schröger, 2007).

2.6. Research questions and hypotheses

The purpose of this study is to investigate neural mechanisms of foreign phoneme acquisition in adulthood and neuroplastic changes related to the learning process using MEG and behavioral measurements. The thesis had four main research questions:

- I. What kind of electromagnetic fields and cortical source currents in the brain are related to phoneme processing and discrimination in minimal pair words?

- II. Data from the pilot study suggested that the discrimination task is rather difficult for untrained subjects. Is it possible to improve foreign phoneme discrimination accuracy in young adults with moderate amount of training?
- III. Does the training affect brain's electromagnetic fields or their source current distributions when processing foreign phonemes in a passive discrimination task? What kind of learning-related neuroplastic changes take place in the brain?
- IV. Is stronger behavioral learning effect associated with greater changes in ERF amplitudes or source current distributions before vs. after training?

Based on previous studies we hypothesized behavioral discrimination accuracy of phonemes III–III to improve at least for a part of Finnish subjects. However, individual differences in learning efficiency were presumed. Training was also expected to affect the neuromagnetic evoked responses. We expected to find stronger ERF amplitudes (MMNm in particular) and source current distribution significances in left superior temporal region after training – at least for subjects who showed behavioral learning effect at some level. Neural activation in the control condition (native phoneme contrast discrimination) was expected to be stronger than in experimental condition and not to change considerably between pre and post-training measurements. Activation levels for novel phoneme categories after training were not expected to reach the level of control phonemes but according to our hypothesis, they should resemble each other more than prior to training.

3. Methods

3.1. Subjects

Data from 20 healthy volunteers (9 women and 11 men) aged between 18–35 (mean 28) was used in this study. In total, 26 volunteers participated in study but due to technical problems six of them had to be excluded from the analysis. All subjects spoke Finnish as their only mother tongue and reported not to have been exposed on foreign languages regularly at home before school age. Moreover, they had no significant previous knowledge of Russian language. Subjects had no diagnosed neurological disorders, learning disabilities or language disorders. All subjects were right-handed according to Edinburgh Handedness Inventory (Oldfield, 1971). Subjects signed an informed consent before every measurement and were compensated for lost working hours and travel expenses. The experiment was approved by the Aalto University Research Ethics Committee.

3.2. Experimental procedure

The experiment consisted of two measurement days and a training period between them during which the subjects practiced ш–ш phoneme discrimination at home by playing a learning game by computer (Figure 2). 1st and 2nd measurement day tasks were identical and consisted of MEG and behavioral discrimination parts. Subjects also filled several background information questionnaires during measurements. After the 2nd measurement day subjects participated in a separate Magnetic Resonance Imaging (MRI) measurement where structural brain images to be used in source modelling were taken. During the last measurement session individual background information was also gathered with questionnaires and psychological tests (these were not included in this thesis).

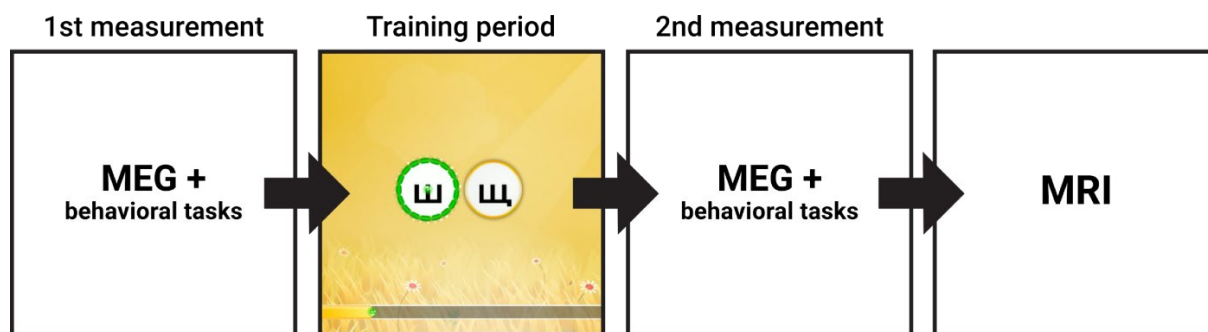


Figure 2. Experimental procedure.

3.3. Stimuli

Stimuli consisted of minimal pairs of Russian pseudowords presented auditorily. We received assistance from a Russian linguist who helped us create pairs of pseudowords that differ from each other only by the first phoneme. That is, each word had two versions that shared the same word body but started with a different phoneme. In the experimental pairs the phoneme alterations were always between ш and щ (e.g. *шкopa–щкopa*; *shkopa–shchkopa*). For control pairs we chose easier Russian phoneme contrasts that are more familiar for Finnish speakers and also exist in Finnish language, at least in some form (e.g. *тaркoн–вaркoн*; *tarkon–varkon*). Used control phoneme contrasts were /g/–/b/, /s/–/n/, /t/–/k/, /z/–/n/, /d/–/p/, /l/–/r/, /t/–/v/ and /v/–/f/. They were chosen so that both phonemes in a pair are associated with two distinct Finnish phoneme categories. Although some of these Russian phoneme categories do not appear in Finnish language as such, they have both 1) distinctive physical characteristics compared to chosen contrast phoneme and 2) either very similar sounding Finnish equivalents (Russian /v/ represents Finnish /v/) or they are commonly used in loanwords (/z/). Both experimental and control pseudowords were created by changing phonemes in actual Russian words. All words were reviewed and checked by a Russian linguist to take Russian phonotactic constraints, transitional probabilities and phonological neighborhood effects into account.

The actual stimuli were produced from digital recordings (44.1 kHz, 16 bit) of native Russian speakers pronouncing both versions of chosen pseudowords. We had four speakers for the measurement stimuli and additional four speakers for the sounds used in the learning game in the training period. In both stimuli sets two of the speakers were female and two were male. Recordings were conducted in an acoustically treated room with Behringer B-2 Pro high-fidelity large diaphragm condenser microphone (Behringer International GmbH, Willich,

Germany) and Adobe Audition CC 2015 recording software (Adobe Systems Inc., San Jose, CA, USA).

3.3.1. Editing

In this experiment we were interested in the acquisition of pure ɪ and ʏ phonemes. Therefore, we wanted to eliminate all differentiation cues other than those related to the spectral and temporal properties of these phonemes. We had two main concerns: unwanted learning effects and MMNm false positives. Firstly, in the pilot experiment we found out that even Russian natives often rely on coarticulation cues of the following vowel when differentiating between these sounds. With some speakers coarticulation almost completely changed the following vowel. Although this is a natural lingual phenomenon, we wanted to exclude it from the experiment as we hypothesized that these vowel changes would be much easier for Finnish people to detect than the actual spectral changes between ɪ and ʏ . Secondly, previous studies have shown that almost any acoustic change, even a small one, in a repetitive sound sequence can trigger MMN component in EEG or MEG signal (e.g. Schröger, 1998). We wanted to eliminate the possibility of false positive MMN components due to uncontrolled physical characteristics of the word body after the phoneme of interest, such as changes in amplitude, pitch, intonation etc. That is, we wanted the only acoustic change within a word pair to be in the changing phoneme. This way we were also able to control unwanted learning effects based on random but repetitive acoustic factors and (both linguistic and paralinguistic) articulatory characteristics in word bodies.

In order to control these factors all stimuli were artificially cross-spliced. Cross-splicing is an auditory stimulus editing technique where certain sound segments are cut and replaced by corresponding segments from another instance of that sound. It has been widely used in neuro-linguistic studies where strict control of acoustic features of the stimuli is required (Steinberg, Truckenbrodt, & Jacobsen, 2012). First, for all speakers we chose the best sounding recording of each word pair with as little acoustic variation as possible. Then we spliced each pair with Adobe Audition CC 2015 software using the ɪ -versions as the basis. That is, we *cross-spliced* all the ɪ -versions by replacing the latter part of the word (e.g. *ɯkɔna*; *shchkɔpa*) with the latter part of corresponding ʏ -version (*ɯkɔna*; *shkɔpa*) leaving the only acoustic change within a spliced pair in the first phoneme (Figure 3). Same cross-splicing procedure was applied to control stimuli also.

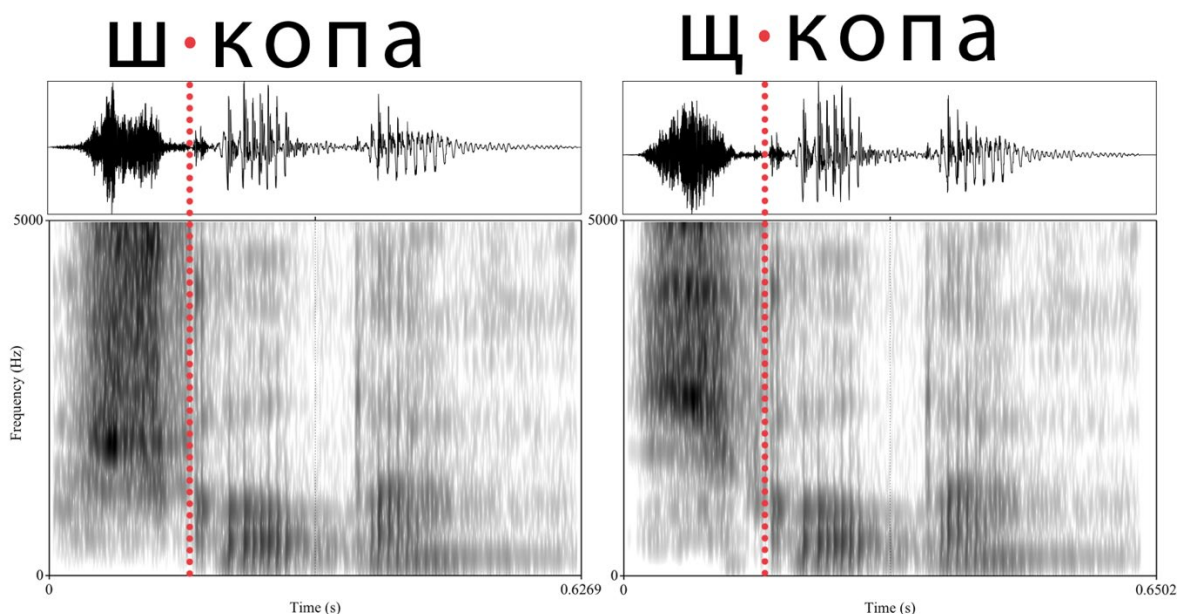


Figure 3. Spectrograms and waveforms of a cross-spliced stimulus pair *шкопа–щкопа*. The latter part of the stimulus is acoustically identical in both versions. *Щкопа* stimulus (on the right) has been artificially edited by replacing the latter part of the word with the corresponding segment of the *шкопа* stimulus (on the left). The left side *ш*-version remains in its originally recorded form.

Щ is usually pronounced a bit longer than ш but we did not want to exclude this temporal cue and artificially standardize phoneme lengths. This caused some variation in stimulus segment lengths. Ш/щ phoneme length in experimental stimuli was 85–280ms and the whole stimulus length 510–1192ms depending on the word and the speaker. All spliced words used in the experiment sounded natural without strange transitions. However, we also did an item analysis based on the piloting data and ended up excluding a few pairs which proved to be exceptionally difficult to discriminate even for a native Russian speaker. Also, a few of the original stimulus items had to be excluded from the experiment due to technical reasons. Number of sibilant and control pairs was balanced in order to maintain same signal-to-noise ratio in the data. In the end we used 18 unique experimental pairs and 18 control pairs. All stimuli were normalized by average intensity and scaled to 72dB using Praat 6.0 (Boersma & Weenink, 2017).

Some researchers have stated splicing should be used with caution as it has been found that artificial splicing of auditory stimuli might produce unwanted interference effects in neural responses compared to natural unedited sounds (Steinberg et al., 2012). We were not too concerned about this since splicing is a widely used technique and most of this critique concerns experimental settings where, for example, splicing related loss of phonological cues interferes with some higher order lingual phenomena. We acknowledge the possible loss of ecological

validity (at least in relation to natural Russian language) but in this experiment, we were interested in general mechanisms of low-level phoneme acquisition and not about any language specific factors or higher order phenomena such as semantics. Therefore, we thought teaching phonemes in a slightly artificial (or “pseudo Russian”) setting was not too big issue in comparison to the benefits of being able to control all stimulus factors and unwanted learning effects.

3.3.2. Training stimuli

In order to avoid familiarity biases the stimuli used in the training period was different from the experimental stimuli and the speakers were different. Learning game sounds included similar pseudowords but also short syllables and separate *iii* and *iii* sounds. Stimuli were recorded in similar conditions and with same equipment that the stimuli used in the measurements. Syllables and pseudowords were also cross-spliced the same way.

3.4. Measurement sessions

1st and 2nd measurements (see Figure 2) were conducted at Aalto University MEG Core at the Department of Neuroscience and Biomedical Engineering. Each measurement session lasted approximately 2.5 hours in total. MEG data were recorded using 306-channel Elekta Neuromag™ MEG device (Elekta Neuromag Oy, Helsinki, Finland). Four electrooculography (EOG) electrodes (plus a ground electrode) were used to record vertical and horizontal eye movements. Five head position indicator (HPI) coils were attached to subjects’ forehead as well as left and right mastoids and digitized with a 3D digitizer. Fiducial markers from three anatomical landmarks and some additional points from scalp surface (to be used in MEG-MRI coregistration) were also digitized for each subject. During the measurements, subjects sat inside a high-end 3-layer magnetically shielded room (Imedco AG, Hägendorf, Switzerland). Before each MEG section subject’s head position, with respect to the sensors, was measured through the HPI coils. In addition, we used continuous head position indication (cHPI) which tracked subjects head movements during the measurements. Nevertheless, the subjects were asked to keep their head as still as possible during MEG sections.

3.4.1. Behavioral discrimination task

Both measurement days started with a behavioral word pair discrimination task during which the subjects sat in the magnetically shielded room, but no MEG data was recorded (yet). In the discrimination task pseudoword pairs were presented auditorily to the subjects. Word pairs contained both minimal pairs (as described before) and pairs where the phoneme stayed the same. The subjects' task was to tell if there were any difference between the two words. The subjects were instructed to pay attention even to the slightest changes in single sounds. They were also aware that the task contained two differing sibilant sounds. 144 experimental (sibilant) pairs and 144 control pairs were presented in randomized order for each subject in both 1st and 2nd measurement. Half of both the experimental and control pairs contained a change in one phoneme and the other half did not. Additionally, in half of these pairs (36 experimental and 36 control pairs) also the speaker changed between the words. Subjects were instructed not to count mere speaker changes as word changes. Stimulus Onset Asynchrony (SOA) between words within a pair was 1700ms (approximately 500ms Inter-Stimulus Interval, ISI) and after the answer there was a 1000ms delay before the next pair. Answer time was not limited but subjects were not able to repeat the pair. Behavioral task was programmed with Presentation® software (Neurobehavioral Systems, Inc., version 20.1).

The task was divided in four sections: two sibilant sound sections (experimental condition) and two control sections (control condition). One section lasted approximately 6–8 minutes and there was a small break between every section. In addition, the subjects were allowed to take short voluntary breaks between every fifth word pair if they felt the need. Control stimuli were presented in separate condition in order to help the subjects to orientate in the actual experimental stimuli (sibilant sounds) which were much more difficult to discriminate than the control pairs. In the pilot study where both stimuli were mixed some subjects did not even try to discriminate the sibilant sounds as their attention was drawn by easier control pairs.

Subjective auditory threshold was measured before the task and the stimuli were presented at 60 dB subjective intensity via a Panphonis SoundShower flat panel speaker. Task instructions were projected in a screen in front of the subjects before the task. Question “*Were the words exactly the same?*” was presented after each pair (in Finnish). Subjects responded with their left and right index finger using an optical answer device. Response hand for ‘yes, same’ and ‘no, different’ was altered between subjects but stayed consistent for single subjects. Answer hand information was visible throughout the task. We wanted to minimize any learning effects

outside the training period, so the subjects did not get feedback on their performance. However, word stimuli in both 1st and 2nd measurements were the same. This might have caused some sort of familiarity effect in the 2nd measurement results but we argue that compared to the learning effect produced by intensive training period this possible familiarity effect is insignificantly small.

3.4.2. MEG oddball task

Behavioral discrimination task was followed by MEG section where same pseudowords were presented in a passive auditory oddball task. In an oddball paradigm multiple sequences of repetitive standard stimuli, interrupted by a deviant stimulus, are presented. In our case the standard stimulus of a stimulus sequence was always one variant of a minimal pair while the other variant acted as a deviant stimulus. Both phonemes *iii* and *iii* were used as standards and deviants alike. In this task the speaker didn't change between a standard and corresponding deviant leaving the only physical difference between a standard and a deviant in the altering phoneme.

36 experimental deviants and 36 control deviants were presented on both measurement days. That is, each word instance (phoneme variants and different speakers included) acted exactly one time as a deviant and one time as a standard on both days. Order of the sequences was randomized and the number of standard stimulus repetitions in a sequence was randomly set between 6–9. Although short inter-stimulus intervals have been associated with stronger MMN responses (Näätänen et al., 2007), we ended up using slightly longer ISI than usual because the shorter ISI tested in the pilot study resulted in poorer auditory responses overall. Therefore, SOA (from onset to onset) was set to 3000ms (ISI approx. 1200ms), which is not exceptional for a regular speech perception study using natural word stimuli, but somewhat longer than in MMN studies typically. We wanted to control SOA over ISI as there was some variation in stimulus lengths and we wanted to maintain a fixed stimulus onset rhythm between standards and deviants throughout different sequences. With this long SOA we decided to keep the number of deviants moderate in order to avoid the task being too long and exhausting for the subjects. Stimulus sequences were programmed with Presentation® software (Neurobehavioral Systems, Inc.).

During the task subjects were sitting in a magnetically shielded room watching a silent Chaplin film from a small projected screen at 140cm distance. During the film they were exposed to

auditory stimuli from the panel speaker (60 dB subjective intensity). Subjects were instructed to concentrate in the silent film and pay no attention to the sounds. However, no follow-up questions were asked about the film, in order to cut down the duration of our already lengthy measurement sessions. Subjects were also told to hold their head still during the measurement and avoid excessive blinking, if possible. The task was divided into two approx. 15-minute sections and the subjects were allowed to have a short break in the middle. Experimental and control sequences were mixed in both sections.

The stimuli used in the oddball task were the same that in the behavioral discrimination task. In order to avoid familiarity effects in behavioral learning results, the behavioral task was always done first. Familiarity could have slightly affected the neural activation starting levels in the 1st measurement oddball task but this should not be a problem when comparing the differences between 2nd and 1st measurements. In the 2nd measurement, however, this “extra training” might indeed have boosted activation by a little but as the effect was similar with every participant this should not be an issue when interpreting the results. Most likely the effect is really small compared to actual training period related learning effects.

3.4.3. MRI session

The 3rd and last research session took place at Aalto University Advanced Magnetic Imaging (AMI) Centre and the adjacent Aalto Behavioral Laboratory (ABL). For each subject structural magnetic resonance (MR) brain images were taken using a 3 T MAGNETOM Skyra whole-body scanner (Siemens Healthcare, Erlangen, Germany), with a standard 20-channel head-neck coil. After the MRI measurement the subjects performed several psychological tests at Aalto Behavioral Laboratory soundproof test room. The subjects also filled additional background information questionnaires concerning their language background, socioeconomic status (SES) and musicality. These tests and questionnaires did not belong to the scope of my thesis and therefore the resulting data was not used or interpreted here.

3.5. Gamified training period

Between 1st and the 2nd measurement subjects practiced III–III phoneme discrimination at home by playing a phoneme learning game on a computer (screenshot in Figure 2). The game was produced by Niilo Mäki Institute in University of Jyväskylä and customized for this experiment

specifically. It is based on GraphoLearn learning game platform (<https://info.grapholearn.com/>) originally designed to help children to learn to read by teaching them letter-sound correspondences of their local language.

The game consisted of three levels: phoneme, syllable and pseudoword level. On each level multiple phonemes, syllables or pseudowords were presented to the subjects via headphones one at a time. Each auditory item presented begun with either m or ʌ sound and subjects were told to categorize it to one or the other group. After each sound the game gave the subjects two clickable choices: “ m ” and “ ʌ ” presented in random order. Instant feedback of correct and incorrect answers was given. On the bottom of the screen a progress bar was shown where the subjects could monitor their performance on current level. No prior information about the phonetical differences between m and ʌ was given and the subjects were told to make their judgements based on trial and error.

One round of the game included all three levels and took approximately 15 minutes to pass. During the training period subjects played 10 rounds of the game which made it roughly two and half hours of active training in total. Subjects were allowed to play two rounds a day at most and but other than that we encouraged them to play as regularly as possible. They were allowed to have intervening days between game rounds but no more than one in a row. We also instructed the subjects to schedule their training so that there would be as few days between the last game round and the 2nd measurement as possible. However, they were not allowed to play on the measurement day. Even though the amount of training was controlled, the length of the training period varied between subjects (typically 7–10 days).

Subjects used their own computers to play the game. They were instructed to find a calm and quiet practicing environment and always use headphones while playing. Game scores and logs were saved in a cloud server where we could monitor that the subjects carried out the exercises accordingly. Thus, permanent Internet connection was required to play the game.

3.6. MEG data analysis

3.6.1. Data & preprocessing

MEG data was recorded with 306 channels (204 gradiometers and 102 magnetometers). As magnetometer data appeared quite noisy, we decided to include only gradiometer data in the

analysis. First, we used temporally-extended signal space separation (tSSS) to suppress signal artifacts coming from sources outside of the brain (Taulu & Hari, 2009). For this we used Elekta Neuromag MaxFilter software which also detected and removed any bad channels from the data. Head movement correction based on cHPI tracking was also done and the data from different measurement days was transformed to the same reference head position for comparable results.

Eye blinks and eye movements produce large artifacts to MEG data. As they overlap with the spherical ‘inside the head’ signal source volume separated by tSSS from ‘outside the head’ sources, they cannot be removed from the data using tSSS (Hari & Puce, 2017). Therefore, we used independent component analysis (ICA) to identify and remove (mostly blink related) ocular artifacts from the data. ICA decomposes the MEG signal to separate components by detecting spatially and temporally distinct signal sources from the data (Makeig & Onton, 2008). Artifact components can be sorted out from brain-related signal sources by comparing the waveforms and topographic maps of the components. Unwanted components can then be eliminated from the data. In addition to manual component evaluation and removal we used automated blink component identification algorithm (Gramfort et al., 2013). First the algorithm identified blink related data segments (epochs) using the EOG channels and then blink related components were identified by analyzing the correlations between the components and these data segments. In order to avoid low-frequency drifts the data were high-pass filtered at 0.1Hz before ICA. A low-pass filter at 40Hz was also applied to avoid possible line-frequency noise. For these processing steps (as well as for most of the rest of the data processing and analyses) MNE-Python software package (versions 0.19.0 & 0.14.1) was used (Gramfort et al., 2014).

3.6.2. Event-related fields

After artifact suppression and filtering the data were segmented to 1200ms event-related epochs (0 to 1000ms from stimulus onset time with a baseline from -200 to 0ms before stimulus onset). With 3000 fT/cm rejection threshold no epochs got rejected. In the oddball task m and m words worked both as standard and as deviant stimuli on their respective turns. In the analysis different types of standards and deviants were merged into single standard and deviant groups regardless of the phoneme. This way we could eliminate possible biasing effects in the MEG responses arising from physical differences between m and m stimuli. In mismatch negativity research this approach is sometimes called identity MMN, or iMMN (e.g. Pulvermüller, Shtyrov, Ilmoniemi, & Marslen-Wilson, 2006). For each subject standards and deviants were averaged

separately for 1st and 2nd measurement day. For averaged standards we included only the last epoch before each deviant. The rest of the items in standard stimulus sequences were ignored. This way the used standard epochs were saturated to the maximum and SNR between averaged standards and deviants stayed the same as the number of items in both groups was the same (Figure 4). MMNm subtraction waveforms (deviant–standard) based on the averaged single subject conditions were also calculated for each subject and measurement day.

Experimental stimuli		Control stimuli	
<i>Before training</i>	<i>After training</i>	<i>Before training</i>	<i>After training</i>
standards (36 epochs)	standards (36 epochs)	standards (36 epochs)	standards (36 epochs)
deviants (36 epochs)	deviants (36 epochs)	deviants (36 epochs)	deviants (36 epochs)
MMNm (deviants – standards)	MMNm (deviants – standards)	MMNm (deviants – standards)	MMNm (deviants – standards)

Figure 4. Single subject conditions in the MEG task.

3.6.3. Source modelling

Preprocessing of MRI data was performed using FreeSurfer 6.0.0 software (Fischl, 2012). Processing steps included motion correction, intensity normalization, removal of non-brain tissue and modelling the cerebral cortex surface. Cortical areas were mapped using HCPMMP1 parcellation (Glasser et al., 2016). Event-related MEG data was then co-registered with corresponding MRI images using MNE-Python. In addition to digitalized anatomical landmarks, coil positions, and other reference points, we had photographs of subjects' left and right ear with pre-auricular points marked to assist with co-registration.

Conductor models used in primary current source estimation were subject-specific and based on individual head geometry. These source spaces were derived from structural MRI images using boundary element method (BEM). Inverse problem was tackled by minimum norm estimation (Gramfort et al., 2013). It is a distributed inverse method based on distributing a large number of dipoles throughout the cortex, estimating their moments, and trying to find the minimum current distribution that explains the MEG activation. Minimum norm estimates (MNE) were calculated for all stimulus conditions and converted to spatio-temporal statistical maps using dynamic statistical parametric mapping, dSPM (Dale et al., 2000). These dSPM maps show event-related prime current distribution significance across cortical surface. To be able to make intersubject comparisons the resulting data from individual subjects was morphed

into a common space. FreeSurfer FsAverage template brain was used in source estimate visualizations.

3.7. Statistical analyses

3.7.1. Behavioral data analysis

Different response strategies might influence the rate of correct answers in the behavioral discrimination task (Macmillan & Creelman, 1990). Therefore, phoneme discrimination efficiency was measured using d' (d-prime) sensitivity index. Coming from signal detection theory this measure takes into account both hit rate (correct answers) and false alarm rate (reporting a difference where there is none) of the responses. It can be used to measure perceptual performance and discriminability in behavioral tasks (Haatveit et al., 2010). The higher the hit rate and the lower the false alarm rate the better target pairs (phonemic change) can be discriminated from nontargets (no change). Consequently, higher d' value indicates more reliable detection of significant acoustic features that separate paired words from each other, and therefore better discrimination ability. Zero value means that the signal cannot be detected from noise at all.

For both measurement days separate d' values were calculated. Learning was considered to have happened if d' value raised along with training. Pre and post training d' values and learning effect (d' change between measurements) were investigated with paired samples t -tests and Pearson's correlation analyses. Behavioral data was analyzed with Python programming language (versions 3.6.9 & 2.7.12).

3.7.2. MEG statistical analysis

When running statistical tests on MEG source estimates the problem of multiple comparisons arises. Testing every data point on the cortex separately raises familywise error rate (the risk of false positive results). Here, multiple comparisons problem was addressed with non-parametric permutation testing. Permutation tests are based on rearranging data point labels randomly multiple times and comparing the results with the experimental condition (Nichols & Hayasaka, 2003). Statistical differences between MNE source estimates in different conditions (see Figure 4) were tested with non-parametric spatio-temporal cluster-based permutation tests (1000 permutations). MEG data typically contains strong spatial and temporal correlations. Therefore

spatio-temporal cluster-based test tries to find contiguous regions where significant activation differences between conditions exist (Gramfort et al., 2013).

Two separate permutation tests with different spatial constraints were used. First the tests were performed for all left hemisphere data without further restrictions (left hemisphere is usually dominant in early language processing thus right hemisphere data was left out from statistical analysis). In addition, tests where analysis was restricted to a specified cortical region of interest (ROI) in the left posterior superior temporal gyrus, were performed. ROI around the left pSTG was composed by splitting and merging cortical areas mapped by multi-modal HCPMMP1 parcellation (Glasser et al., 2016). All tests were restricted to 100ms time windows around brain responses of interest. Same time windows were used for all subjects and conditions. The exact windows were determined by calculating the temporal center of mass of MNE source current responses (mean of all subjects and conditions). The grand averaged data revealed only one clearly defined neuromagnetic response which peaked at 208ms (around the expected MMNm latency). Thus, only one time window at 158–258ms was used (see results for further information).

4. Results

4.1. Behavioral learning results

Behavioral phoneme discrimination efficiency in minimal pair task was measured using d' sensitivity index. Higher d' score indicates better discrimination ability.

There was a slight increase in d' mean (all subjects) from 1st to 2nd measurement (Figure 5). The d' mean before training was 0.923 (SD = 0.612) and after training 1.202 (SD = 0.644). The difference was not statistically significant ($t = -1.37$, $p = 0.186$) yet individual differences in learning efficiency were considerable. Altogether 13 subjects showed improved d' score after training (d' change ranging from 0.137 to 1.742; mean = 0.802, SD = 0.603) indicating at least some sort of training-related learning. For the rest 7 subjects d' change was negative (ranging from -0.208 to -1.371; mean = -0.691, SD = 0.367) and no learning effect was found. It seemed that in the latter ‘non-learner’ group training might even have made some subjects perform worse in the discrimination task (Figure 5).

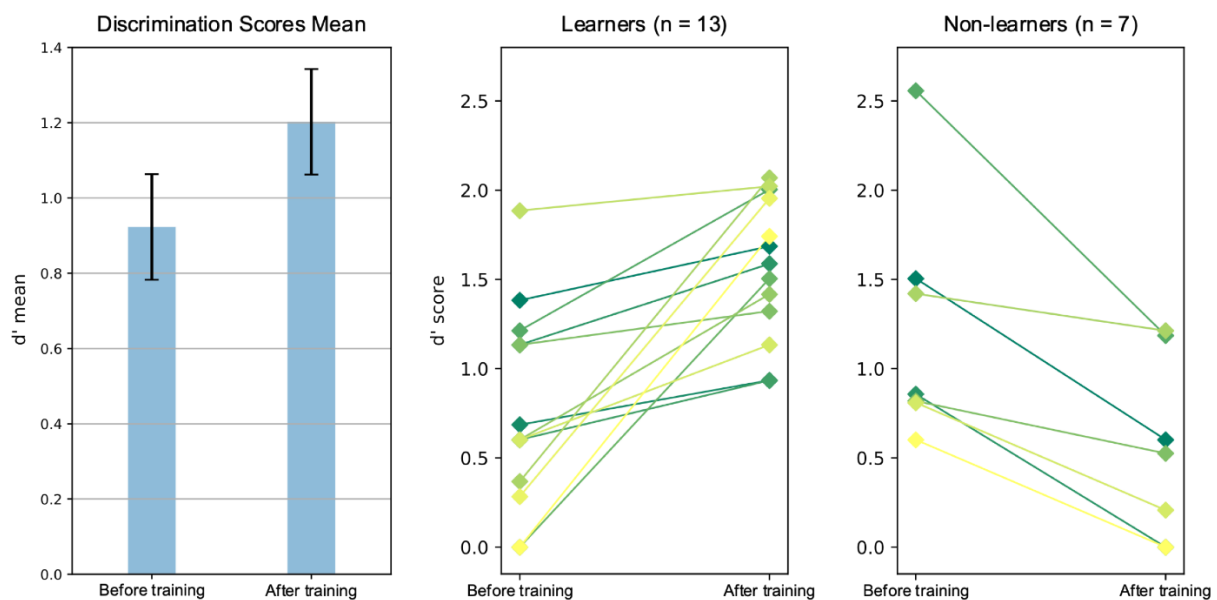


Figure 5. Behavioral discrimination ability scores for experimental stimuli before and after training. The d' mean (on the leftmost chart) increased from 0.923 to 1.202 (black bars represent the standard error of the mean) but the difference was not statistically significant ($t = -1.37$, $p = 0.186$). Individual differences in learning are plotted on the right. 13 subjects improved their d' score whereas 7 subjects performed worse after training.

Correlation analysis revealed a negative correlation between d' score before training (starting level) and d' change between the measurements ($r = -0.689$, $p < 0.001$). This suggests that

lower discrimination ability prior to training was connected to stronger training-related learning effect.

Discrimination scores for control stimuli were considerably better. The d' in the first measurement was 3.075 (SD = 0.412) and 3.517 (SD = 0.513) in the second. The performance was better in the second measurement ($t = -3.562$, $p = 0.002$) even though control phonemes were not practiced during the training period.

4.2. MEG sensory ERFs

For almost all subjects both standard and deviant stimuli elicited a clear neuromagnetic event-related field around 200ms in the left temporal channels (Figure 6). Later increase in activation between 300–800ms was also found. The 200ms response had higher peak amplitudes and was temporally better defined and more consistent across subjects and different conditions. Later responses, on the other hand, were flatter, more scattered and varied considerably in amplitude and latency across subjects and conditions. No single activation peak could be identified hence late activation was left out from further statistical evaluation.

On average, there were no considerable amplitude or latency differences between standards and deviants in the 200ms response. ERF amplitudes for deviants seemed slightly stronger than standards in both measurements but the difference was very small (sensory data were not statistically evaluated, though). MMNm waveforms (deviant–standard) were calculated but no clear MMNm responses could be identified based on the sensory data (Figure 6). However, visual inspection of subjects' data revealed a substantial amount of individual variability. For some subjects standard and deviant waveforms were almost identical but for others deviant response amplitudes were either higher or lower than standard amplitudes. Amplitude differences tended to be greater in the 1st measurement trials whereas 2nd measurement mean amplitudes for the standards often approached mean amplitudes for the deviants. On average 200ms ERF amplitudes for deviants seemed slightly stronger than standards in both measurements.

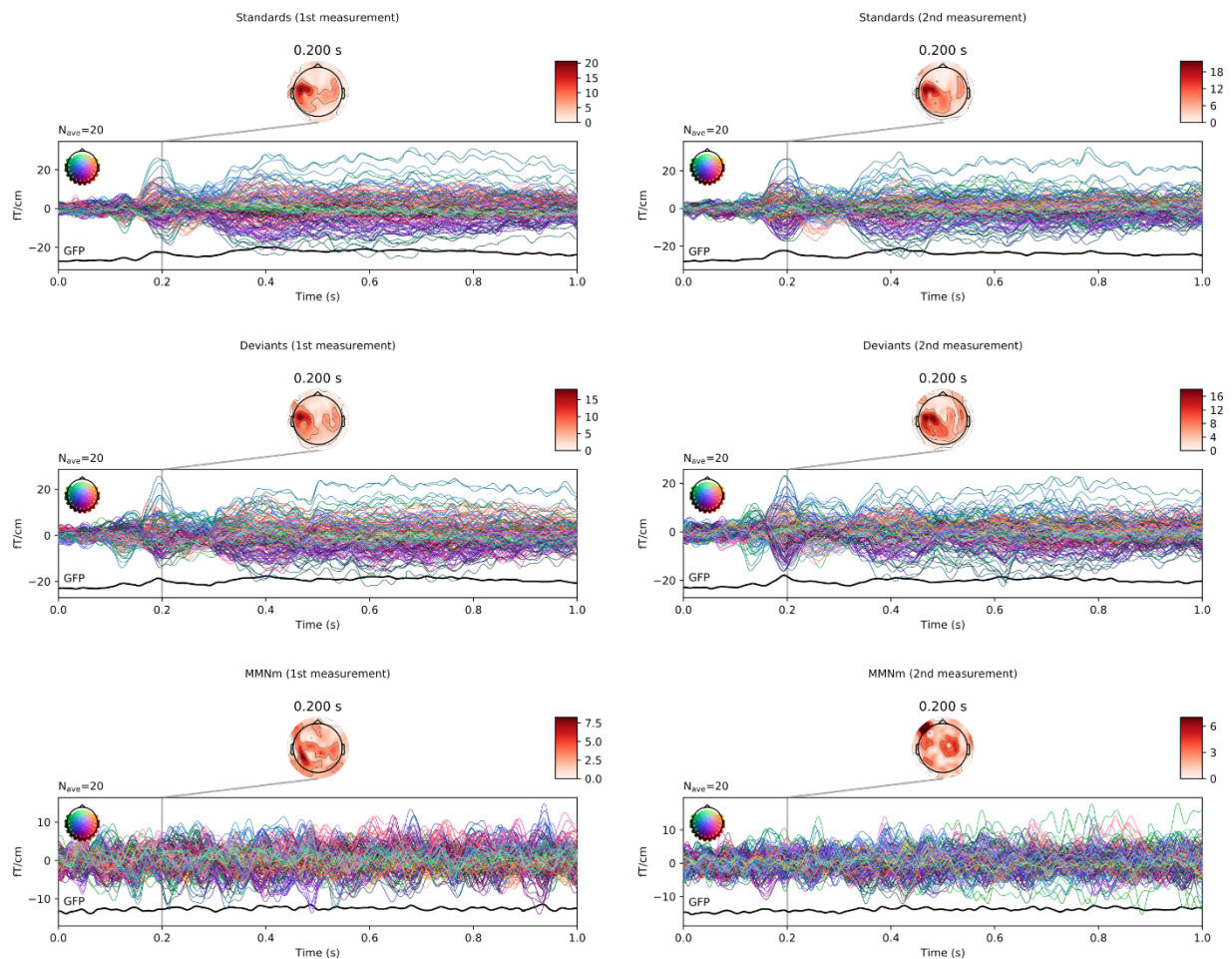


Figure 6. Grand averaged sensory ERFs for experimental stimuli before (on the left side) and after training (on the right). MEG gradiometer channels are plotted in butterfly charts and color coded to show the sensor positions. Global Field Potential (GFP) is plotted below. Topographic maps show the spatial distribution of the magnetic fields at the 200ms activation peak. Activation for the 200ms ERF is strongest in the left temporal sensors. Later increase in activation can be seen after 300ms but it is more dispersed and lacks a recognizable activation peak. Differences in peak amplitudes between standard and deviants were very small in both measurements, though deviant responses look slightly stronger on average.

Similar sensory responses in left temporal channels were found for control stimuli (Figure 7). Unexpectedly, no distinct MMNm responses were found for control stimuli either. Surprisingly, the 200ms ERF amplitudes seemed stronger for standard rather than deviant stimuli in both measurements, but the difference was again very subtle. The later activation behaved more coherently in control than experimental conditions and appeared little later, after 400ms. Average activation peak was around 600ms.

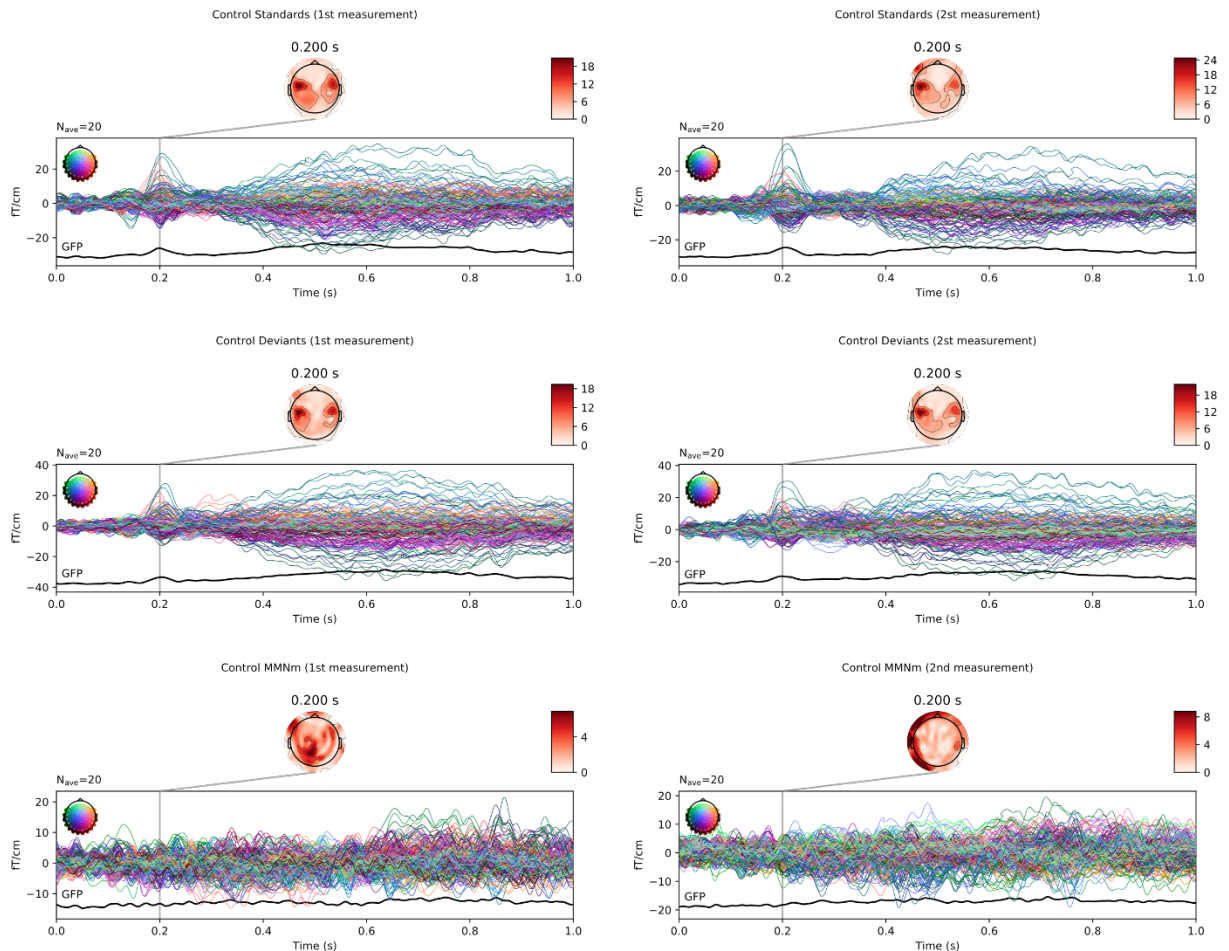


Figure 7. Grand averaged sensory ERFs for control stimuli before (on the left side) and after training (on the right). MEG gradiometer channels are plotted in butterfly charts and color coded to show the sensor positions. Global Field Potential (GFP) is plotted below. Topographic maps show the spatial distribution of the magnetic fields at the 200ms activation peak. Activation for the 200ms ERF is strongest in the left temporal sensors. Later increase in activation can be seen after 400ms peaking around 600ms. Differences in the 200ms ERF peak amplitudes between standard and deviants were very small in both measurements, though standard responses look slightly stronger on average.

4.3. MEG source modelling

Primary current distribution of event-related MEG data was estimated using minimum norm estimates (MNE) and dynamic statistical parameter mapping (dSPM). Source current distribution maps (the mean of all subjects) showed stronger statistical significance for currents in the left than right hemisphere. For both standard and deviant stimuli dSPM significance was strongest in the left temporal brain areas; more specifically in the posterior superior temporal gyrus (pSTG). Current distribution for the 200ms ERF response was more focal whereas later activation was spatially and temporally more scattered. Source modelling produced similar results for control stimuli conditions. Average center of the mass for the earlier ERF was found

208ms post stimulus at $X = -55.12$, $Y = -28.11$, $Z = 3.86$ (MNI coordinates). Based on this, 100ms time window (158–258ms) was used in statistical tests. Average temporal center of mass for the later activation was found 600ms post stimulus but since the activation was widely scattered and there was no clear activation peak, I did not find it reasonable to perform statistical tests in the later time window.

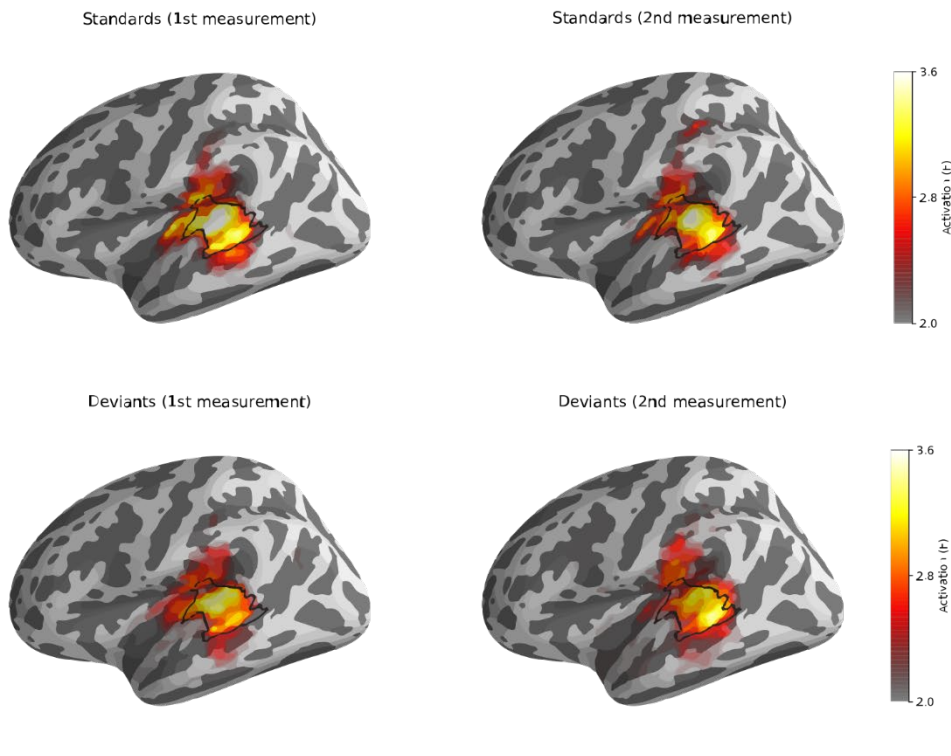
Differences in primary current distributions (dSPM maps) between different conditions were tested using permutation tests based on spatio-temporal clustering. Two separate tests were conducted: a left hemisphere test including all possible left hemisphere sources, and ROI test restricted to the selected region of interest in left hemisphere pSTG.

4.3.1. MMNm responses

Visual inspection did not reveal notable differences between primary current distributions of standard and deviant stimuli in experimental conditions on either of the measurement days (Figure 8). Left hemisphere permutation test findings were coinciding: no statistically significant differences were found between standard and deviant dSPM distribution. ROI-restricted permutation test provided similar results. Since no significant changes in activation were found, it appears that no MMNm responses were elicited in experimental conditions on sample-level.

Control stimuli dSPM maps showed marginal (yet statistically insignificant) differences between standard and deviant conditions in early time window (Figure 8). On the first measurement day neuromagnetic evoked responses for deviants seemed stronger than standards. The opposite was true after training: standards appeared to be stronger than deviants. However, neither left hemisphere nor ROI restricted permutation tests could confirm these observations. In the late time window after 300ms deviant responses seemed stronger at least before training. Deviant activation seemed to decrease slightly with training. No statistical tests were carried out for later responses to confirm the observations.

Experimental stimuli



Control stimuli

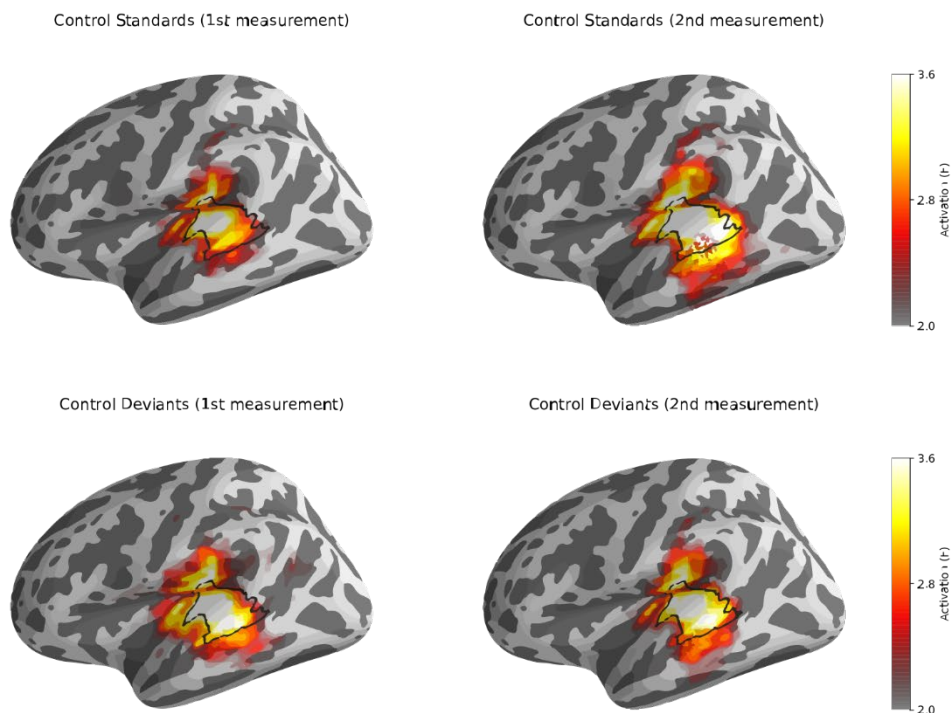


Figure 8. Average source current distribution maps (dSPM) in the left hemisphere for the early response (208ms after stimulus onset). Data are morphed to a template brain where the cortex is shown inflated (darker areas = gyri, lighter areas = sulci). Selected ROI around the left pSTG is circulated by a black line. Activation before training is plotted on the left and after training on the right. No statistically significant differences in activation were found between standards and deviants in any of the conditions. No significant training-related changes were found either. Overall activation for control stimuli seemed somewhat stronger than for experimental stimuli.

4.3.2. Training-related effects

Training-related effects were studied by comparing 1st measurement results with 2nd measurement results. Since no clear MMNm responses were found, changes in MMNm (deviant–standard) source current significances were not studied.

Visual inspection did not reveal notable differences between primary current distributions before and after training in experimental conditions in the early time window (Figure 8). Also, based on both left hemisphere and ROI-restricted permutation tests, current distributions of 2nd measurement conditions did not significantly differ from 1st measurement conditions. That is, no apparent training-related changes in the earlier ERF source estimates were found in conditions averaged over all subjects. However, in the late time window dSPM significance seemed to decrease slightly after training for both standard and deviant stimuli. Again, the difference in late activation was not statistically studied, though.

On control condition permutation tests did not reveal significant differences between 1st and 2nd measurement. However, activation for deviant stimuli seemed to decrease slightly whereas activation for standards increased with training in the early time window (Figure 8). Also, in the later time window activation for deviants seemed to decrease along with training.

4.4. Interactions between MEG and behavioral results

In order to visualize the interactions between MEG and behavioral results, source current distribution maps were grouped and averaged based on the direction of the behavioral learning effect. Learner group (n = 13) consisted of subjects that showed improved d' score after training. Non-learner subjects (n = 7) on the other hand had lower d' scores in the 2nd measurement than before training.

Results for the learner group are plotted in Figure 9. In general dSPM activation significance in the early time window seemed to increase along with training by a little for both standards and deviants (although the difference might not be evident in the 208ms snapshot of the source current distribution map in Figure 9). There were hardly noticeable differences between standards and deviants before training, yet after training a relatively small difference could be

observed. On average deviant responses seemed to be slightly stronger and spatially more scattered after training. The difference was not statistically significant, though ($p > 0.05$).

Learner group

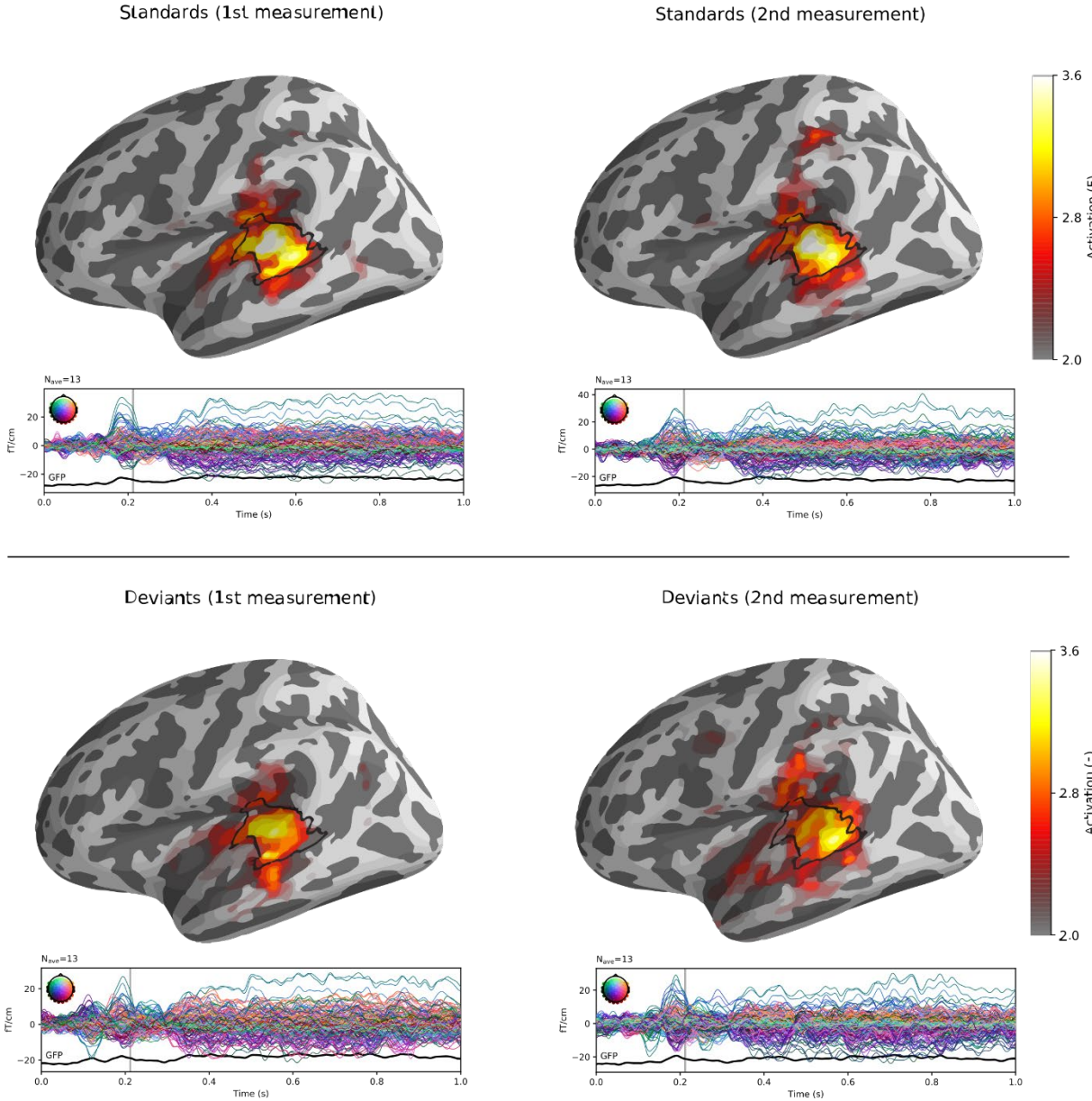


Figure 9. Average source current distribution dSPM maps for the learner group ($n = 13$) in the left hemisphere 208ms after stimulus onset (average temporal center of the mass). Corresponding sensory data are plotted below (208ms time point is indicated by a gray vertical line). Activation before training is plotted on the left and after training on the right. A subtle statistically insignificant training-related increase in activation for the deviants could be observed.

Non-learner group showed slightly different results (Figure 10). Activation for both standards and deviants seemed to decrease a bit by training. Interestingly though, the effect seemed

slightly stronger for the deviant stimuli than for standards. No noticeable differences between standards and deviants were observed. Overall the activation was spatially more scattered in the non-learner group than in the learners, especially before training. Nevertheless, all the differences were again very small and statistically insignificant.

Non-learner group

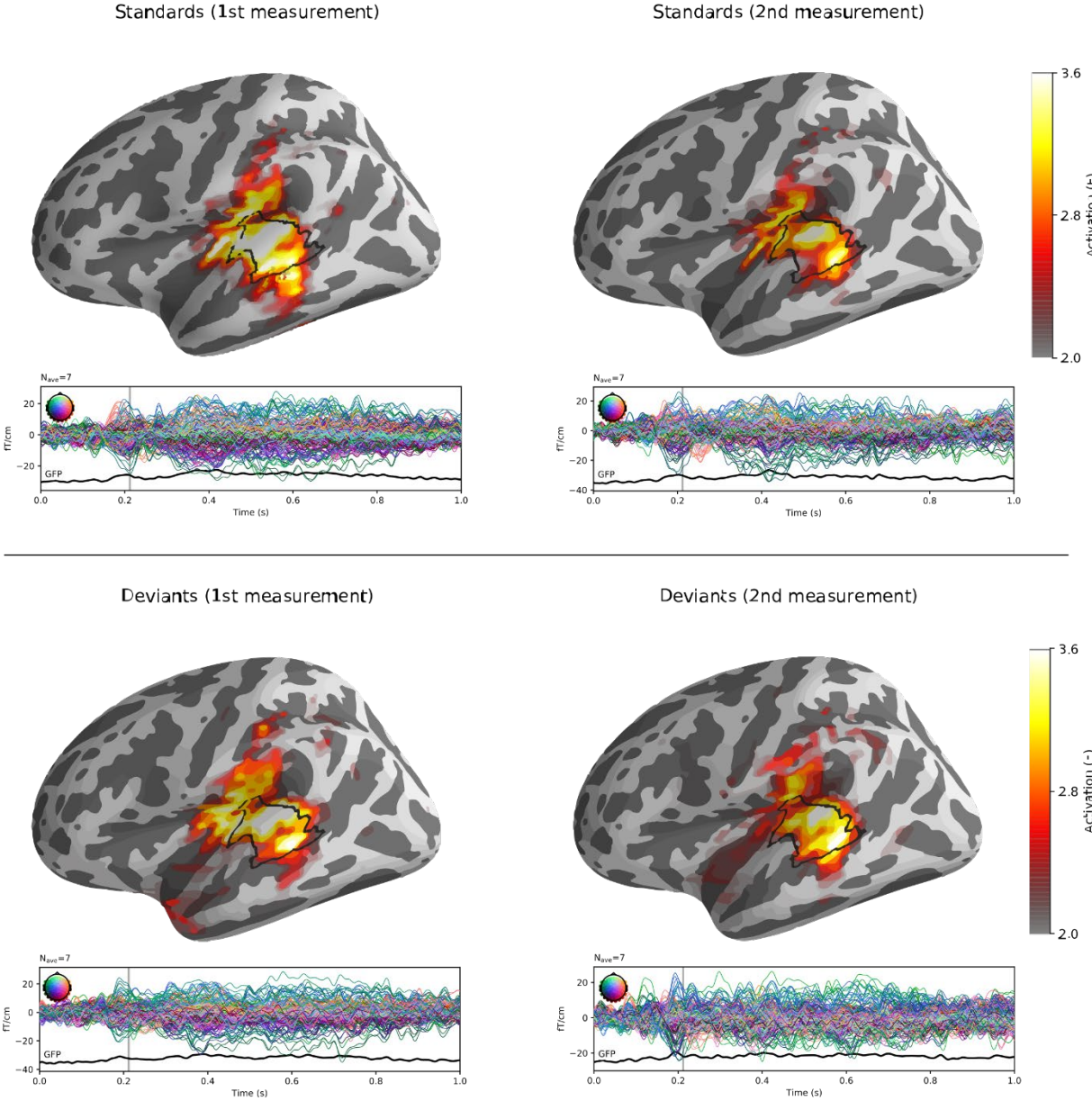


Figure 10. Average source current distribution dSPM maps for the non-learner group ($n = 7$) in the left hemisphere 208ms after stimulus onset (average temporal center of the mass). Corresponding sensory data are plotted below (208ms time point is indicated by a gray vertical line). Activation before training is plotted on the left and after training on the right. Activation seemed to decrease by training, yet the differences were statistically insignificant.

Sensory ERFs plotted below the dSPM maps in Figure 9 and Figure 10 show that the average temporal center of the mass for the early response (gray vertical line) is a few milliseconds later than the ERF amplitude peaks. Both early and late ERFs are stronger and better defined for learners than non-learners. For learners the activation is strongest in the left temporal sensors whereas for non-learners it seems to lean more towards the left frontal sensors.

Interestingly, control condition results seemed to be somewhat opposite to experimental condition. For learners, dSPM activation for control deviants before training was stronger than for standards. Activation significance for standards grew along with training whereas for deviants it decreased, and in the second measurement standards appear to be stronger than deviants. Similar effect was seen on sample-level but for learners it was more pronounced. Non-learner group did not show as clear differences in control condition as learners. Nevertheless, activation for standards seemed to decrease slightly along with training while deviant activation increased. Activation was more focal after training than before. Overall activation levels were lower with non-learners than learners. However, again, none of these observations had any statistical significance.

5. Discussion

This study examined the neural mechanisms of foreign language phoneme learning in adulthood. Naïve Finnish subjects were trained to discriminate foreign Russian phonemes ш /ʃ/ and щ /ɕ(:)/ which are acoustically very much alike. Phoneme discrimination ability was measured before and after training both behaviorally and with MEG (oddball task). The main hypotheses were that training would improve behavioral discrimination accuracy of the two phonemes and result in stronger MMNm responses in the brain (particularly in the pSTG) during passive discrimination task.

Behavioral discrimination ability did not improve by training as much as expected (no statistically significant learning effect on sample-level), yet individual differences in learning were considerable. Altogether 13 of 20 subjects improved their d' score post training. As expected, neuromagnetic evoked responses for standard and deviant stimuli peaking approximately at 200ms and located around the pSTG were found. However, contrary to the hypotheses, the differences between standard and deviant responses were very subtle, thus no clear MMNm responses could be found either before or after training. A slight training-related increase in source current amplitude for the deviant stimuli could be observed in the learner group around 200ms in the pSTG area. The effect was opposite for control stimuli and did not exist in the non-learner group. Although small, the effect was in accordance with the hypotheses, suggesting some sort of change in neural processing in the learner group. Still, none of the observed differences were statistically significant.

5.1. Behavioral learning results

Subjects' phoneme discrimination ability increased slightly along with training, but the effect was not statistically significant (Figure 5). Although it is likely that with a larger sample size statistically significant results could be obtained, we were expecting a stronger learning effect. However, individual differences in learning were considerable. 13 subjects improved their performance (average d' change 0.802) and 7 subjects performed worse after training (average d' change -0.691). Subjects were considerably better discriminating control phonemes, as expected. Interestingly, while training did not significantly improve discrimination

performance for experimental phonemes, for control stimuli it did. Since no control phonemes were included in the learning game the effect is most likely due to the familiarity of the task and the stimuli. It has been found that automatic perceptual learning of novel spoken words can take place rapidly even with minor perceptual exposure (Kimppa, Kujala, Leminen, Vainio, & Shtyrov, 2015).

A significant correlation between learning and the starting level of discrimination ability was also found. Lower discrimination ability prior to training was connected to stronger training-related learning effect ($r = -0.689$, $p < 0.001$). Specifically, the strongest learning effects were found on those subjects that scored very low in the first measurement. After training these subjects often reached the average discrimination ability levels. On the other hand, the subjects who performed on average level already in the first measurement often showed minimal improvement in discrimination level after training, or none at all. It is possible that our short training paradigm was not enough to properly facilitate learning beyond the basic level of discrimination, offering high-scoring subjects little room for improvement. It is also possible that some subjects underachieved in the first measurement for some reason (e.g. poor answering technique, inexperienced ear or misunderstanding task instructions). For these subjects the considerable improvement in performance in the second measurement could therefore, to a certain extent, be explained by the familiarity of the task. Strong correlation between starting level and learning effect can also partly (but not wholly) be explained by one outlier subject who performed extremely good in the first measurement and very poorly in the second measurement.

Results from the behavioral task were poorer than expected. This implies that the training paradigm used in this study was not optimal for the chosen set of stimuli. Both the amount and quality of training should be assessed carefully in future studies when training subjects with difficult foreign phoneme contrasts.

5.2. MEG results

5.2.1. MMNm responses

Clear neuromagnetic event-related fields for standard and deviant stimuli were found peaking around 200ms in the left temporal channels (Figure 6). This falls into the expected MMNm time window (usually peaking at 150–250ms after stimulus onset). Neural source currents for

the 200ms response were located around the pSTG as expected (Figure 8). The exact temporal center of the mass of was found at 208ms after stimulus onset which was a little later than the observed ERF peaks. However, no significant amplitude or latency differences in sensory MEG data were found between standards and deviants in the 200ms ERF. Therefore, contrary to the hypotheses, no MMNm responses could be interpreted based on the sensory MEG data. No significant differences were found in source current distributions either. Training had no significant effect neither on experimental stimuli ERFs nor their source current distributions.

One probable reason for the insignificance of the MMNm responses is that the acoustic difference between standard and deviant stimuli was too subtle. Chosen phoneme contrast was difficult for the subjects as the overall performance in the behavioral task indicates. Yet, it is curious that no significant MMNm responses were found in control condition either, even though performance in behavioral discrimination task with control stimuli was quite good. Minor changes between standards and deviants in control condition could be visually observed, though. Before training activation for deviant seemed slightly stronger than for standards. This might be a pure coincidence, however, since the observed changes were not statistically significant. The activation difference between standards and deviants also turned the other way around after training which is in contradiction to the hypotheses.

5.2.2. Relationship between MEG results and learning

In order to analyze interactions between behavioral and MEG results subjects were divided to learner and non-learner groups based on their performance in the behavioral task. The original hypothesis was that stronger learning effect would result in stronger MMNm responses. No significant MMNm responses for experimental stimuli on sensory or source level were found on either learner or non-learner group (Figure 9 and 10). However, minor statistically insignificant differences in source current distributions between standards and deviants supporting the hypothesis could be observed. For learners, deviant responses seemed to grow slightly stronger after training compared to standard responses (although not to the expected extent). Similar effect was not present in the non-learner group. Instead, for non-learners training seemed to strengthen standard rather than deviant responses.

Source estimation results from control condition proved to be even more interesting, even though, once again, no statistically significant effects were found. For learners, similar but more pronounced effect was found as in sample-level. Activation strength for deviants exceeded the

level of standards before training as expected. However, after training activation for standards was stronger than for deviants. This might have something to do with the brain adapting to familiar deviant stimuli, as same stimuli were used both measurements. Interestingly though, similar effect was not present in the non-learner group.

Evidence from both behavioral and MEG results indicates that either the amount or quality of training was not sufficient for the subjects to learning the chosen phoneme contrast. Or the other way around: the artificially cross-spliced Russian phoneme contrast in itself was too difficult for Finnish subjects to acquire in such a short time. Though tiny and statistically insignificant, the observed training-related effects in the learner group were in line with the hypotheses. If the overall behavioral learning effect was stronger the training-related changes in the brain most likely would have been more significant.

5.2.3. Late activation

In addition to the earlier ERF around 200ms a later increase in activation was also observed. Average temporal center of mass for the later activation was found 600ms after stimulus onset. However, later responses were flatter, their source current distributions more scattered and overall variability between different subjects and conditions was considerable. Hence later activity was left out from statistical analysis. Interestingly, on grand average level late activation seemed to decrease by training. This could be linked to stimulus or task familiarity.

Although the oddball task was passive and subjects were supposedly attending to the film, later activity probably reflects some sort of higher-level cognitive processing. One possibility is that it, at least to some extent, reflects so called *late negative difference* (late Nd) usually found in dichotic listening or other dual tasks. It is thought to be linked to selective attention processes after stimulus identification (Singhal, Doerfling, & Fowler, 2002). Although stronger late Nd responses are associated with attended rather than unattended stimuli, the fact that late Nd is reduced by practice is line with the observations in current study. Inter-stimulus interval used in the oddball task was also quite long and stronger late Nd is associated with longer ISI (Singhal et al., 2002). However, since individual differences in later responses were considerable, the activity probably consists of several overlapping components.

5.3. Methodological limitations

Various issues and methodological choices in current study can be identified that might also explain why the results turned out not as unambiguous as expected. Some of these questions and possible challenges (especially those concerning the size of the behavioral learning effect) raised already after the piloting phase of the study. Even though we knew that the learning task was very difficult, and the overall learning effect might turn out to be small, we did not want to change the design of the experiment. We had already invested in the customized learning game and making great changes in such late phase of the project would have resulted in additional costs and a major delay in the project. The other reason was that a research question concentrating on individual differences in learning efficiency based on the same dataset was investigated elsewhere. Therefore, variability in subjects' performance and a weaker overall learning effect were not seen as a major problem.

5.3.1. Stimuli

The phoneme contrast used in this study was carefully selected. The aim was to find a contrast that would be very difficult for untrained Finnish subjects to hear. The hypothesis was that almost any phoneme contrasts, no matter how difficult, could be learned with sufficient amount of practice. Since the aim was to study the general language-independent phoneme acquisition mechanism in adulthood, several languages, even artificially created phonemes, were considered to be used the study. Natural language stimuli felt like the best option in terms of ecological validity and motivating the subjects to properly engage in the learning process. The decision to use Russian language was both practical and based on its distinct phonological characteristics compared to Finnish. Russian linguist was consulted both in the selection of the phonemes and the creation of stimulus pseudowords to ensure the linguistic validity of the paradigm.

Because MMNm response is elicited by practically any acoustic change, the stimuli were artificially edited and controlled using cross-splicing method. Since we wanted to teach the subjects to perceive the actual spectral differences between the phonemes all unwanted categorizing cues were cut out from the stimuli. (Temporal cues [phoneme length] were not eliminated but they were marginal compared to spectral differences.) However, unexpectedly weak learning results in current study suggest that spectral cues might not play as big part in phoneme categorization than we originally thought. Especially when differentiating between

phonemes with strong acoustic resemblance other cues, such as coarticulation, context (phonological or semantic) and sound length, might be more important for categorization even for native speakers. This of course depends on the language and phonemes at question. The two phonemes selected in this study are exceptionally closely related by their spectral features (Figure 1) which is why other categorization cues are likely emphasized. Although cross-splicing is a widely used method, it has been also criticized. For example, Steinberg et al. (2012) found an MMN response for natural German syllables but failed to replicate the results with cross-spliced stimuli because of missing coarticulation cues. Using natural stimuli in this experiment might have led to more significant behavioral and MEG results. (On the other hand, it would have made interpreting the results even more difficult.)

Several subjects reported that articulatory differences between speakers made learning the phonemes very hard in the learning game. It is possible that some speakers had slightly different manners of speech or a dialect of some sort. Also, training stimuli was spoken by different speakers than the stimuli used in the measurements. If speakers' individual differences in pronunciation were significant, the differentiation skills subjects acquired during training might not have transferred to the measurement task to their full extent.

One of the reasons why no clear MMNm responses were found might have to do with the nature of sibilant sounds. Sibilants and other fricatives do not have as well-defined voice onset times as affricate consonants that begin with full stop. Therefore, the resulting auditory brain responses might be temporally more dispersed and flatter, thus being more difficult to interpret. An important question is then why the control stimuli, which contained also affricate consonants, did not behave as anticipated in the MEG task. The phoneme contrasts used in the control condition were familiar for the subjects and the categorization performance in the behavioral task was at the expected level, yet no clear MMNm responses were detected. There are several possible explanations. Firstly, control stimuli included various different consonant phoneme contrasts (both fricatives and affricates). Different phonemes vary by their acoustic features which might result in differences in latencies, amplitudes and the overall form of elicited brain responses. When averaged such differences flatten and disperse the resulting mean responses making it more difficult to interpret whether MMNm components exist in the data. Secondly, several speakers were used when recording the stimuli. The idea was to increase the ecological validity the study but at the same time individual differences in articulation probably produced unwanted variance in the auditory responses. Using only one speaker and

selecting a single well-controlled phoneme contrast as the control stimuli might have fixed the issue, although loss of ecological validity would have been inevitable.

The lack of evident MMNm responses in both experimental and control condition could also be explained simply by the fact that the physical differences between standard and deviant stimuli were too small. Because of cross-splicing the only difference between standard and deviant word pairs lied in the first phoneme (Figure 3). The length of the phoneme of interest varied between 85–280ms but the length of the whole stimulus could be up to 1192ms. Since most of the stimulus timeline was physically identical between a standard and a deviant, the overall differences within stimulus pairs were marginal. Auditory responses from the rest of the stimulus word are also likely to have interfered with the responses related to the change in the first phoneme. Thanks to cross-splicing, single standard and deviant pairs were meticulously controlled. However, the length of the starting phoneme nor the onset and length of the rest of the word were controlled between different stimulus words and speakers. This might have resulted in misalignments, smearing of the effect, overlapping auditory responses and other unwanted interference effects during averaging.

5.3.2. Number of repetitions

Even though the sample size ($n = 20$) was satisfying, the number of stimulus repetitions especially in the MEG tasks was relatively low (36 deviants per condition). The decision was made mostly to save time in the measurement and make it more bearable for the subjects. In hindsight it can be argued that the number of deviant trials was too low for this complex experimental setting. Even though certain robust ERP/ERF components have been found with as low as 10–15 trials, the statistical power of the results increases enormously as the number of repetitions grows higher (Boudewyn, Luck, Farrens, & Kappenman, 2018). In fact, Duncan et al. (2009) recommend as much as 150 deviant trials in a standard auditory MMN experiment using artificial non-speech stimuli. Although the number of trials is rarely that high in studies that use more complex natural stimuli, this might still be the single most important reason why no statistically significant results were found in this experiment.

The pilot results suggested using longer 3000ms SOA (ISI approx. 1200ms) than originally planned. Auditory brain responses with a shorter SOA were quite poor and we suspected it might be due to overlapping responses. 1200ms ISI is not overly long (Duncan et al., 2009, recommend using 500–1000ms ISI in auditory MMN studies). However, since the stimuli

themselves were quite lengthy, the measurement session was quite long for the subjects even with this few trials. Longer ISI has been associated with weaker MMN responses (Näätänen et al., 2007) which probably also affected the results. Since the effects found in MEG data were small, using shorter SOA and increasing the number of deviant trials could have drastically increased the SNR and statistical significance of the results making smaller effects pop up.

5.3.3. Time windows of interest

The delay of the time window used in the statistical tests was determined by the temporal center of mass of grand averaged source estimate data. Using grand averaged data to select the time windows for statistical tests has been criticized by some researchers because it increases the risk of type II error and false findings (Luck & Gaspelin, 2017). Even though not the most elegant, this is not an uncommon approach in neuroscience studies. Since the length of the time window as well as the spatial region of interest were selected a priori, we argue that the risk of biased results is negligible. Also, in this case as no significant results were acquired this concern can be ignored without harm.

5.4. Conclusions and future directions

In this study Finnish subjects were trying to learn the difference between Russian phonemes ш /ʃ/ and ш /ɕ(:)/. Previous studies (e.g. Näätänen et al., 1997; Tremblay et al., 1997; Winkler et al., 1999; Ylinen et al., 2009) have shown that adult brain can learn to differentiate even subtle acoustic differences in foreign or artificial speech sounds. In preceding EEG and MEG studies training has resulted in stronger MMN or MMNm responses in oddball-type tasks (e.g. Menning et al., 2002; Tremblay et al., 1997, 1998; Ylinen et al., 2009). In this experiment we failed to find as strong effects as the evidence from previous studies suggested. Individual differences were considerable, but no statistically significant group-level changes were found neither on the behavioral task nor the MEG responses. More than half of the subjects improved their behavioral discrimination accuracy after approximately one week of training, but the effect was weaker than expected. A slight training-related increase in source current amplitude for deviant stimuli was observed in the learner group around 200ms in the pSTG area. Even though statistically insignificant, the effect was in accordance to the hypotheses, suggesting some sort of change in neural processing in the learner group.

It is likely that methodological limitations affected the results and decreased the effect size in both behavioral and MEG tasks. Based on observations from the data and existing research evidence it is plausible that the chosen foreign phoneme contrast could be properly acquired with sufficient amount of training. In future studies the required amount of training should be more carefully evaluated and matched to the difficulty of the stimuli. Cross-splicing as a stimulus controlling method should be used with caution especially when studying acoustic differences that are very small and hard to perceive. Finally, and most importantly, the number of required stimulus repetitions in the MEG task should be re-evaluated and the amount of trials per condition increased respectively.

Despite the possible methodological limitations, current study was not able to replicate the findings from various previous studies (e.g. Näätänen et al., 1997; Tremblay et al., 1997, 1998; Winkler et al., 1999; Ylinen et al., 2009). It is, therefore, possible that acquiring difficult foreign phonemes after childhood years might not be as easy as some of the studies have proposed, or at least it is dependent of the nature of the phoneme contrast in question. Most of the previous research on phonemic learning in adulthood has concentrated on vocals or simple duration cues (e.g. Menning et al., 2002; Näätänen et al., 1997; Tremblay et al., 1997; Winkler et al., 1999; Ylinen et al., 2009). The generalizability of these findings on learning to differentiate between closely related fricatives or other consonants based novel spectral cues can therefore be questioned. More research on a wider spectrum of different phoneme types and differentiation cues is needed in order to get a more comprehensive understanding on how the general phoneme acquisition mechanism in brain functions.

Our results also suggest that there are major individual differences in learning efficiency when trying to acquire difficult unfamiliar phoneme contrasts in adulthood. Based on the difficulty of the stimuli and the variance in the results we suggest that individual differences in, for example, subjects' background might play a larger role in learning efficiency when the differentiation task is more demanding. As to which background variables affect phonemic learning, and to what extent, will be left for future studies to investigate. Bearing this in mind, we already collected information about subjects' socioeconomic, educational, linguistic and musical background as well as their personality and intelligence to be analyzed in another research paper. The interactions between behavioral and neuromagnetic data will also be re-analyzed using a regression model rather than simple learner–non-learner grouping. Examining individual differences in foreign phoneme acquisition is important, as it helps to better

understand the biological mechanisms and environmental factors that affect foreign language learning in general, offering valuable insight for educational facilities, for instance.

6. References

- Aaltonen, O., Eerola, O., Hellström, Å., Uusipaikka, E., & Lang, A. H. (1997). Perceptual magnet effect in the light of behavioral and psychophysiological data. *The Journal of the Acoustical Society of America*, *101*(2), 1090–1105. <https://doi.org/10.1121/1.418031>
- Birdsong, D. (2005). *Interpreting Age Effects in Second Language Acquisition. Handbook of bilingualism: Psycholinguistic approaches.*
- Boersma, P., & Weenink, D. (2017). Praat: doing phonetics by computer. Retrieved from <http://www.praat.org/>
- Boudewyn, M. A., Luck, S. J., Farrens, J. L., & Kappenman, E. S. (2018). How Many Trials Does It Take to Get a Significant ERP Effect? It Depends. *Psychophysiology*, *55*(6).
- Buchsbaum, B. R., Hickok, G., & Humphries, C. (2001). Cognitive Science: A Multidisciplinary Role of left posterior superior temporal gyrus in phonological processing for speech perception and production. *Cognitive Science*, *25*(784375790), 663–678. <https://doi.org/10.1207/s15516709cog2505>
- Chang, E. F., Rieger, J. W., Johnson, K., Berger, M. S., Barbaro, N. M., & Knight, R. T. (2010). Categorical Speech Representation in Human Superior Temporal Gyrus. *Nature Neuroscience*, *13*(11), 1428–1432. <https://doi.org/10.1038/nn.2641>
- Cheour, M., Ceponiene, R., Lehtokoski, A., Luuk, A., Allik, J., Alho, K., & Näätänen, R. (1998). Development of language-specific phoneme representations in the infant brain. *Nature Neuroscience*, *1*(5), 351–353. <https://doi.org/10.1038/1561>
- Dale, A. M., Liu, A. K., Fischl, B. R., Buckner, R. L., Belliveau, J. W., Lewine, J. D., & Halgren, E. (2000). Dynamic statistical parametric mapping: Combining fMRI and MEG for high-resolution imaging of cortical activity. *Neuron*, *26*(1), 55–67. [https://doi.org/10.1016/S0896-6273\(00\)81138-1](https://doi.org/10.1016/S0896-6273(00)81138-1)
- Dehaene-Lambertz, G., Dupoux, E., & Gout, A. (2000). Electrophysiological correlates of phonological processing: A cross-linguistic study. *Journal of Cognitive Neuroscience*, *12*(4), 635–647. <https://doi.org/10.1162/089892900562390>
- Duncan, C. C., Barry, R. J., Connolly, J. F., Fischer, C., Michie, P. T., Näätänen, R., ... Van Petten, C. (2009). Event-related potentials in clinical research: Guidelines for eliciting, recording, and quantifying mismatch negativity, P300, and N400. *Clinical Neurophysiology*, *120*(11), 1883–1908. <https://doi.org/10.1016/j.clinph.2009.07.045>
- Fischl, B. (2012). FreeSurfer. *NeuroImage*, *62*(2), 774–781.

<https://doi.org/10.1016/j.neuroimage.2012.01.021>

- Glasser, M. F., Coalson, T. S., Robinson, E. C., Hacker, C. D., Harwell, J., Yacoub, E., ... Van Essen, D. C. (2016). A multi-modal parcellation of human cerebral cortex. *Nature*, *536*(7615), 171–178. <https://doi.org/10.1038/nature18933>
- Gramfort, A., Luessi, M., Larson, E., Engemann, D. A., Strohmeier, D., Brodbeck, C., ... Hämäläinen, M. (2013). MEG and EEG data analysis with MNE-Python. *Frontiers in Neuroscience*, *7*(7 DEC), 1–13. <https://doi.org/10.3389/fnins.2013.00267>
- Gramfort, A., Luessi, M., Larson, E., Engemann, D. A., Strohmeier, D., Brodbeck, C., ... Hämäläinen, M. S. (2014). MNE software for processing MEG and EEG data. *NeuroImage*, *86*, 446–460. <https://doi.org/10.1016/j.neuroimage.2013.10.027>
- Haatveit, B. C., Sundet, K., Hugdahl, K., Ueland, T., Melle, I., & Andreassen, O. A. (2010). The validity of d prime as a working memory index: Results from the Bergen n-back task. *Journal of Clinical and Experimental Neuropsychology*, *32*(8), 871–880. <https://doi.org/10.1080/13803391003596421>
- Hakulinen, A., Vilkuna, M., Korhonen, R., Koivisto, V., Heinonen, T., & Alho, I. (2004). *Iso suomen kielioppi*. Helsinki: Suomalaisen Kirjallisuuden Seura.
- Hämäläinen, M., Hari, R., Ilmoniemi, R. J., Knuutila, J., & Lounasmaa, O. V. (1993). Magnetoencephalography – theory, instrumentation, and applications to noninvasive studies of the working human brain. *Reviews of Modern Physics*, *65*(2).
- Hari, R., & Puce, A. (2017). *MEG–EEG Primer*. New York: Oxford University Press.
- Helmholtz, H. (1853). Ueber einige Gesetze der Vertheilung elektrischer Ströme in körperlichen Leitern mit Anwendung auf thierisch-elektrischen Versuche. *Annalen Der Physik*, *165*(6), 211–233.
- Imada, T., Zhang, Y., Cheour, M., Taulu, S., Ahonen, A., & Kuhl, P. K. (2006). Infant speech perception activates Broca’s area: A developmental magnetoencephalography study. *NeuroReport*, *17*(10), 957–962. <https://doi.org/10.1097/01.wnr.0000223387.51704.89>
- Kimppa, L., Kujala, T., Leminen, A., Vainio, M., & Shtyrov, Y. (2015). Rapid and automatic speech-specific learning mechanism in human neocortex. *NeuroImage*, *118*, 282–291. <https://doi.org/10.1016/j.neuroimage.2015.05.098>
- Knecht, S., Deppe, M., Dräger, B., Bobe, L., Lohmann, H., Ringelstein, E. B., & Henningsen, H. (2000). Language lateralization in healthy right-handers. *Brain*, *123*(1), 74–81. <https://doi.org/10.1093/brain/123.1.74>
- Kraus, N., Carrell, T. D., King, C., Tremblay, K., & Nkol, T. (1995). Central Auditory System Plasticity Associated with Speech Discrimination Training. *Journal of Cognitive*

- Neuroscience*, 7(1), 25–32.
- Kuhl, Patricia K. (1991). Human adults and human infants show a “perceptual magnet effect” for the prototypes of speech categories, monkeys do not. *Perception & Psychophysics*, 50(2), 93–107.
- Kuhl, Patricia K. (2010). Brain Mechanisms in Early Language Acquisition. *Neuron*, 67(5), 713–727. <https://doi.org/10.1016/j.neuron.2010.08.038>
- Kuhl, Patricia K., Williams, K. A., Lacerda, F., Stevens, K. N., & Lindblom, B. (1992). Linguistic experience alters phonetic perception in infants by 6 months of age. *Science*, 255(5044), 606–608. <https://doi.org/10.1126/science.1736364>
- Kujala, T., & Näätänen, R. (2010). The adaptive brain: A neurophysiological perspective. *Progress in Neurobiology*, 91(1), 55–67. <https://doi.org/10.1016/j.pneurobio.2010.01.006>
- Kujala, T., Tervaniemi, M., & Schröger, E. (2007). The mismatch negativity in cognitive and clinical neuroscience: Theoretical and methodological considerations. *Biological Psychology*, 74(1), 1–19. <https://doi.org/10.1016/j.biopsycho.2006.06.001>
- Luck, S. J. (2014). *An introduction to the event-related potential technique*. MIT press.
- Luck, Steven J., & Gaspelin, N. (2017). How to get statistically significant effects in any ERP experiment (and why you shouldn't). *Psychophysiology*, 54(1), 146–157. <https://doi.org/10.1111/psyp.12639>
- Macmillan, N. A., & Creelman, C. D. (1990). Response Bias: Characteristics of Detection Theory, Threshold Theory, and “Nonparametric” Indexes. *Psychological Bulletin*, 107(3), 401–413. <https://doi.org/10.1037/0033-2909.107.3.401>
- Maddieson, I. (1984). *Patterns of sounds*. Cambridge University Press.
- Makeig, S., & Onton, J. (2008). ERP Features and EEG Dynamics. *Oxford Handbook of Event-Related Potential Components*, 1–53. <https://doi.org/10.1093/oxfordhb/9780195374148.013.0035>
- Menning, H., Imaizumi, S., Zwitserlood, P., & Pantev, C. (2002). Plasticity of the human auditory cortex induced by discrimination learning of non-native, mora-timed contrasts of the Japanese language. *Learning and Memory*, 9(5), 253–267. <https://doi.org/10.1101/lm.49402>
- Mesgarani, N., Cheung, C., Johnson, K., & Chang, E. F. (2014). Phonetic feature encoding in human superior temporal gyrus. *Science*, 343(6174), 1006–1010. <https://doi.org/10.1126/science.1245994>
- Miyawaki, K., Jenkins, J. J., Strange, W., Liberman, A. M., Verbrugge, R., & Fujimura, O. (1975). An effect of linguistic experience: The discrimination of [r] and [l] by native

- speakers of Japanese and English. *Perception & Psychophysics*, 18(5), 331–340.
<https://doi.org/10.3758/BF03211209>
- Näätänen, R., Paavilainen, P., Rinne, T., & Alho, K. (2007). The mismatch negativity (MMN) in basic research of central auditory processing: A review. *Clinical Neurophysiology*, 118(12), 2544–2590. <https://doi.org/10.1016/j.clinph.2007.04.026>
- Näätänen, Risto, Lehtokoski, A., Lennes, M., Cheour, M., Huotilainen, M., Iivonen, A., ... Alho, K. (1997). Language-specific phoneme representations revealed by electric and magnetic brain responses.
- Nichols, T., & Hayasaka, S. (2003). Controlling the familywise error rate in functional neuroimaging: A comparative review. *Statistical Methods in Medical Research*, 12(5), 419–446. <https://doi.org/10.1191/0962280203sm341ra>
- Oldfield, R. C. (1971). The Assessment and Analysis of Handedness: The Edinburgh Inventory. *Neuropsychologia*, 9, 97–113.
- Padgett, J., & Zygis, M. (2007). The evolution of sibilants in Polish and Russian. *Journal of Slavic Linguistics*, 15(2), 291–324.
- Phillips, C., Marantz, A., McGinnis, M., Pesetsky, D., Wexler, K., Yellin, E., ... Rowley, H. (1995). Brain mechanisms of speech perception. *MIT Working Papers in Linguistics*, 26.
- Pulvermüller, F., Shtyrov, Y., Ilmoniemi, R. J., & Marslen-Wilson, W. D. (2006). Tracking speech comprehension in space and time. *NeuroImage*, 31(3), 1297–1305. <https://doi.org/10.1016/j.neuroimage.2006.01.030>
- Schröger, E. (1998). Measurement and interpretation of the mismatch negativity. *Behavior Research Methods, Instruments, & Computers*, 30(1), 131–145. <https://doi.org/10.3758/BF03209423>
- Singh, K. D., Holliday, I. E., Furlong, P. L., & Harding, G. F. A. (1997). Evaluation of MRI-MEG/EEG co-registration strategies using Monte Carlo simulation. *Electroencephalography and Clinical Neurophysiology*, 102(2), 81–85. [https://doi.org/10.1016/S0921-884X\(96\)96570-4](https://doi.org/10.1016/S0921-884X(96)96570-4)
- Singhal, A., Doerfling, P., & Fowler, B. (2002). Effects of a dual task on the N100-P200 complex and the early and late Nd attention waveforms. *Psychophysiology*, 39(2), 236–245. <https://doi.org/10.1017/S0048577202011009>
- Singleton, D. (2001). Age and Second Language Acquisition. *Annual Review of Applied Linguistics*, 21(77).
- Steinberg, J., Truckenbrodt, H., & Jacobsen, T. (2012). The role of stimulus cross-splicing in an event-related potentials study . Misleading formant transitions hinder automatic

- phonological processing. *The Journal of the Acoustical Society of America*, 131(3120).
<https://doi.org/10.1121/1.3688515>
- Taulu, S., & Hari, R. (2009). Removal of magnetoencephalographic artifacts with temporal signal-space separation: demonstration with single-trial auditory-evoked responses. *Human Brain Mapping*, 30(5), 1524–1534. <https://doi.org/10.1002/hbm.20627>
- Tervaniemi, M., Medvedev, S. V., Alho, K., Pakhomov, S. V., Roudas, M. S., Van Zuijen, T. L., & Näätänen, R. (2000). Lateralized automatic auditory processing of phonetic versus musical information: A PET study. *Human Brain Mapping*, 10(2), 74–79. [https://doi.org/10.1002/\(SICI\)1097-0193\(200006\)10:2<74::AID-HBM30>3.0.CO;2-2](https://doi.org/10.1002/(SICI)1097-0193(200006)10:2<74::AID-HBM30>3.0.CO;2-2)
- Tremblay, K., Kraus, N., Carrell, T. D., & McGee, T. (1997). Central auditory system plasticity: Generalization to novel stimuli following listening training. *The Journal of the Acoustical Society of America*, 102(6), 3762–3773. <https://doi.org/10.1121/1.420139>
- Tremblay, K., Kraus, N., & McGee, T. (1998). The time course of auditory perceptual learning: Neurophysiological changes during speech-sound training. *Neuro Report*, 9, 3557–3560.
- Vyshedskiy, A., Mahapatra, S., & Dunn, R. (2017). Linguistically deprived children: meta-analysis of published research underlines the importance of early syntactic language use for normal brain development. *Research Ideas and Outcomes*, 3, e20696. <https://doi.org/10.3897/rio.3.e20696>
- Werker, J. F., & Tees, R. C. (1984). Cross-Language Speech Perception: Evidence for Perceptual Reorganization During the First Year of Life. *Infant Behavior and Development*, 7, 49–63.
- Winkler, I., Kujala, T., Tiitinen, H., Sivonen, P., Alku, P., Lehtokoski, A., ... Näätänen, R. (1999). Brain responses reveal the learning of foreign language phonemes. *Psychophysiology*, 36(5), 638–642. <https://doi.org/10.1017/S0048577299981908>
- Ylinen, S., Uther, M., Latvala, A., Vepsäläinen, S., Iverson, P., Akahane-yamada, R., & Näätänen, R. (2009). Training the Brain to Weight Speech Cues Differently: A Study of Finnish Second-language Users of English, 1319–1332.
- Zygis, M. (2003). Phonetic and phonological aspects of Slavic sibilant fricatives. *ZAS Papers in Linguistics*.