Audiovisual integration of words and processing of native and non-native speech sounds

Jost, Lea B

Posted at the Zurich Open Repository and Archive, University of Zurich
ZORA URL: https://doi.org/10.5167/uzh-123889
Published Version

Originally published at:
Jost, Lea B. Audiovisual integration of words and processing of native and non-native speech sounds. 2015, University of Zurich, Faculty of Arts.
Audiovisual Integration of Words and Processing of Native and Non-Native Speech Sounds

Thesis
(cumulative thesis)
Presented to the Faculty of Arts and Social Sciences
of the University of Zurich
for the Degree of Doctor of Philosophy

by Lea Jost

Accepted in the Spring Term 2015
on the recommendation of the Doctoral Committee:
Prof. Dr. Urs Maurer (main advisor)
Prof. Dr. Klaus Oberauer
Prof. Dr. Volker Dellwo

Zurich, 2015
Für…

…meine Eltern Pia und Hans-Peter, die uneingeschränkt an mich glauben

…meinen Freund Marc, mit dem das Leben doppelt schmeckt

…und für Frank, meinen kleinen – grossen Bruder
# TABLE OF CONTENT

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>SUMMARY</td>
<td>6</td>
</tr>
<tr>
<td>ZUSAMMENFASSUNG</td>
<td>7</td>
</tr>
<tr>
<td>ABBREVIATIONS</td>
<td>8</td>
</tr>
<tr>
<td>1. GENERAL INTRODUCTION</td>
<td>9</td>
</tr>
<tr>
<td>1.1 Audiovisual integration</td>
<td>11</td>
</tr>
<tr>
<td>1.2 Phonological processing</td>
<td>12</td>
</tr>
<tr>
<td>1.3 Methodological aspects of EEG/ERP processing</td>
<td>13</td>
</tr>
<tr>
<td>1.3.1 Topographic ERP approach</td>
<td>13</td>
</tr>
<tr>
<td>1.3.2 Effects of filtering</td>
<td>14</td>
</tr>
<tr>
<td>1.4 Conclusion</td>
<td>15</td>
</tr>
<tr>
<td>2. INTEGRATION OF SPOKEN AND WRITTEN WORDS IN BEGINNING READERS: A</td>
<td>17</td>
</tr>
<tr>
<td>TOPOGRAPHIC ERP STUDY</td>
<td></td>
</tr>
<tr>
<td>3. NATIVE AND NON-NATIVE SPEECH SOUND PROCESSING AND THE NEURAL</td>
<td>18</td>
</tr>
<tr>
<td>MISMATCH RESPONSES: A LONGITUDINAL STUDY ON CLASSROOM-BASED FOREIGN</td>
<td></td>
</tr>
<tr>
<td>LANGUAGE LEARNING</td>
<td></td>
</tr>
<tr>
<td>4. GENERAL DISCUSSION</td>
<td>19</td>
</tr>
<tr>
<td>4.1 Audiovisual integration</td>
<td>19</td>
</tr>
<tr>
<td>4.2 Phonological processing</td>
<td>20</td>
</tr>
<tr>
<td>4.3 Methodological aspects of EEG/ERP processing</td>
<td>23</td>
</tr>
<tr>
<td>4.4 Limitations and outlook</td>
<td>24</td>
</tr>
<tr>
<td>REFERENCES</td>
<td>28</td>
</tr>
<tr>
<td>ACKNOWLEDGEMENTS</td>
<td>44</td>
</tr>
<tr>
<td>CURRICULUM VITAE</td>
<td></td>
</tr>
</tbody>
</table>
Summary

Recent development of modern neuroscience technologies allows investigating processing of language and its neural markers in a non-invasive way.

Study 1 investigated audiovisual integration of familiar and unfamiliar words. Data-driven topographic EEG analyses revealed that audiovisual integration takes place in children who had only 1 year of reading training. Moreover, audiovisual stimuli result in a suppressive neural response compared to the summated unimodal response. Furthermore, audiovisually matching words elicit neural activation in different brain regions compared to audiovisually nonmatching words. Presumably due to lexical-semantic processing, the latter effect was only present for familiar, but not for unfamiliar words.

Study 2 investigated the effects of classroom-based foreign language learning on the processing of native and non-native speech sounds. Results showed that children’s mismatch negativity (MMN) was larger for native than non-native speech sound deviance. Moreover, the limited intensity of one year of classroom-based learning does not seem to lead to changes in processing non-native speech sounds. Furthermore, for the first time a new filter analysis was applied on children’s mismatch response data, clarifying some inconsistencies in the field of developmental MMN research.
Zusammenfassung

Moderne neurowissenschaftliche Technologie ermöglicht es, die Verarbeitung von Sprache und deren neuronalen Korrelate auf nicht-invasive Weise zu erforschen.


In **Studie 2** wurden die Effekte des Fremdsprachenlernens in der Schule auf die Verarbeitung muttersprachlicher und nicht-muttersprachlicher Sprachlaute untersucht. Die Ergebnisse zeigten, dass die MMN (mismatch negativity) bei Kindern als Reaktion auf eine muttersprachliche Abweichung grösser war als bei einer nicht-muttersprachlichen Abweichung. Zudem scheint die begrenzte Intensität von einem Jahr Fremdsprachenlernen in der Schule zu keinen Veränderungen in der Verarbeitung nicht-muttersprachlicher Sprachlaute zu führen. Eine zum ersten mal angewandte Filteranalyse auf die 'Mismatch Response'-Daten von Kindern konnte zudem einige Unbeständigkeit auf dem Gebiet der MMN Forschung aufklären.
### Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>3D</td>
<td>3-dimensional</td>
</tr>
<tr>
<td>ANOVA</td>
<td>analysis of variance</td>
</tr>
<tr>
<td>EEG</td>
<td>electroencephalogram</td>
</tr>
<tr>
<td>ERP</td>
<td>event related potential</td>
</tr>
<tr>
<td>fMRI</td>
<td>functional magnetic resonance imaging</td>
</tr>
<tr>
<td>GFP</td>
<td>global field power</td>
</tr>
<tr>
<td>ISI</td>
<td>interstimulus interval</td>
</tr>
<tr>
<td>LDN</td>
<td>late deviant negativity</td>
</tr>
<tr>
<td>LORETA</td>
<td>low resolution electromagnetic tomography</td>
</tr>
<tr>
<td>MEG</td>
<td>magnetoencephalogram</td>
</tr>
<tr>
<td>MMN</td>
<td>mismatch negativity</td>
</tr>
<tr>
<td>MMR</td>
<td>mismatch response</td>
</tr>
<tr>
<td>PET</td>
<td>positron emission tomography</td>
</tr>
<tr>
<td>PT</td>
<td>planum temporale</td>
</tr>
<tr>
<td>RT</td>
<td>reaction time</td>
</tr>
<tr>
<td>SOA</td>
<td>stimulus analysis of variance</td>
</tr>
<tr>
<td>STS</td>
<td>superior temporal sulcus</td>
</tr>
<tr>
<td>STG</td>
<td>superior temporal gyrus</td>
</tr>
<tr>
<td>TANOVA</td>
<td>topographic analysis of variance</td>
</tr>
</tbody>
</table>
1. General introduction

Not only does language involve talking, listening and writing, but it is also the key to thought and action, communication and interaction. It is thus not surprising, that language has become a popular topic of interest for many researchers from a wide range of fields. One issue that research has focused on is foreign language learning. Some studies investigating foreign language learning in natural and immersive settings have shown that the earlier a new language is acquired, the more proficient one becomes in this language in terms of pronunciation (e.g., Flege et al., 1999), as well as morphosyntax (e.g. Johnson and Newport, 1989). Together with the notion that age of acquisition seems to be one of the key predictors for foreign language learning success, also globalization, multiculturalism and multilingualism have led to the introduction of early foreign language teaching already in elementary school in many countries, including Switzerland. In the canton of Zurich, English is being taught from 2nd grade onwards. Conducting studies early in primary school thus allows investigating different aspects of mother tongue acquisition, but also effects of foreign language learning both on the neural as well as the behavioral level.

Young children learn their mother tongue effortlessly and rapidly. They learn from experience, implicitly and informally (for a review see Kuhl & Rivera-Gaxiola, 2008). However, when children enter school and start learning to read and write, they learn to decode print and to link graphemes to corresponding phonemes, in an explicit and formal manner (Abutalebi et al., 2007). Most children acquire these letter-sound correspondences surprisingly easy (Blomert, 2002), even though our brains are not phylogenetically equipped to learn written language (van Atteveldt et al., 2004). Explicit instruction is necessary in order to integrate audiovisual objects such as letter-speech pairs, which is why attaining fluency in reading takes several years (e.g. Vaessen and Blomert, 2010). Integration of letters and corresponding speech sounds is a long-lasting process which is thus very different from learning spoken language, i.e. audiovisual speech, which involves the integration of natural audiovisual stimuli (for a review see Blomert and Froyen, 2010).
Next to the acquisition of reading and writing skills, another long-lasting process characterized by its explicit and formal manner is classroom-based foreign language learning. Typically, 2-5 lectures per week are devoted to the foreign language in a classroom-based setting (e.g., Muñoz, 2008). This limited exposure contrasts immersion education, where at least half of the subjects are taught in the foreign language (e.g., Lyster and Genesee, 2012), providing the means to learn a target language naturally and implicitly. Behavioral studies indicate that foreign language learning is particularly successful in young children (e.g., Flege et al., 1999). However, there are several studies indicating that age of acquisition is not the only factor, but that socio-economic status (e.g., Nikolov, 2009), psychological factors (e.g., motivation, affective state, beliefs and preferences; Benson, 2013) and in particular quantity and quality of exposure (e.g., Muñoz, 2008; 2011; Pfenninger, 2014), also contribute to the success in foreign language learning.

Modern neuroscience techniques such as fMRI (functional magnetic resonance imaging), PET (positron emission tomography), and EEG (electroencephalogram) allow investigating neural correlates of specific cognitive functions and how they are related to neural processes and circuits in the brain. The use of EEG allows analyzing differences in the size, distribution and timing of neural activity across the scalp for different experimental conditions (e.g., Luck, 2005, Handy, 2005). Importantly, it provides information about sensory, motor, perceptual and cognitive processes in milliseconds (e.g., Kutas and Federmeier, 2000). By providing this exceptionally high temporal resolution, EEG is particularly suited to investigate the time course of language processing. The field of language is extremely diverse and the processes involved very complex. Even though knowledge about its structure, function and neurobiological basis has increased a lot over the past decades, there are still many open questions regarding the neural basis of language (Friederici, 2011). Subject to many studies are for example phonological processing and audiovisual integration. These two processes are both involved in spoken and written language and are highly connected to each other. It has been hypothesized, that learning to read changes phonological representations by forming new letter-sound associations (Blomert and Froyen, 2010). Importantly, phonological processing and audiovisual integration are involved in mother tongue acquisition and learning to read, but also when it
comes to foreign language learning. In the following, both phonological processing and audiovisual integration will be addressed and experimental paradigms suited for investigating these processes introduced.

1.1 Audiovisual integration

A first step in reading acquisition is setting up grapheme-phoneme associations and thus linking print to corresponding spoken language (e.g., Ehri, 2005 for a review). Hence, integrating visual and auditory information is a key element of reading (e.g., Blomert, 2011). Neuroimaging studies applying audiovisual paradigms provide critical insights towards understanding of auditory-visual associations. Specifically, they unravel if, where and when audiovisual integration takes place. Typically, an audiovisual experiment consists of an auditory-only condition (A), a visual-only condition (V) and an audiovisual condition (AV). In the case of the audiovisual condition, the stimuli presented auditorily and visually can either be the same, i.e. matching (AVM), or different, i.e. non-matching (AVN).

Two main approaches have been applied in several neuroimaging studies in order to investigate audiovisual integration. One of the approaches contrasts the sum of the unimodal auditory and visual ERPs (event related potentials) to the audiovisual matching condition (SumA+V vs AVM; Raij et al., 2000; Besle et al., 2004; Giard and Peronnet, 1999). The assumption is that in case the response to the bimodal condition differs from the response to the sum of the unimodal conditions, this would be evidence for cross-modal interaction (Besle et al., 2004, additive model). The other approach contrasts the audiovisual matching condition to the audiovisual non-matching condition (AVM vs AVN; Blau et al., 2010; van Atteveldt et al., 2004). This approach postulates that unisensory inputs need to be integrated successfully in order to get different responses for matching and non-matching information (van Atteveldt et al., 2007a,b).

Studies investigating letter-speech sound associations using spatially sensitive methods showed that the (left) superior temporal sulcus (STS) plays a crucial role for audiovisual integration of natural associations (e.g. audiovisual speech), but also for more arbitrary letter-speech sound
associations (van Atteveldt et al., 2004; Raij et al., 2000). Moreover, when the unimodal stimuli are presented simultaneously, feedback mechanisms from the STS to the auditory cortex seem to be involved (van Atteveldt et al., 2007a; see also Blomert and Froyen, 2010 for a review). However, the precise network involved in multimodal integration also seems to be dependent on the experimental paradigms that are being used (see Calvert, 2001 for a review).

1.2 Phonological processing

Phonological processing plays an important role in mapping auditory information onto higher levels of language processing. Moreover, it is also involved in temporarily storing verbal information in working memory (Burton, 2001). Whereas phonology describes how sounds function within a given language or across languages, the branch of phonetics is concerned with the physical properties of speech sounds, i.e. their acoustic and physiological properties and their perception (e.g. Ladefoged and Johnson, 2014). Research on phonetic perception has shown that young infants have the capacity to differentiate between phonetic contrasts of all languages, but that this capacity diminishes between the 6th and the 12th month of age (e.g., Kuhl et al., 2006). When native language phonetic abilities increase (e.g., Cheour et al, 1998; Kuhl et al., 1992; Sundara et al., 2006; for a review see Kuhl and Rivera-Gaxiola, 2008), the brain gets primed to the main language, hence reducing perceptual abilities for non-native languages (Cheour et al., 1998; Best and Roberts 2003, Kuhl et al., 2006).

A neurophysiological brain response recorded in the EEG and MEG (magnetoencephalogram) that has been shown to be suitable for studying phonological specialization is the mismatch negativity (MMN, Näätänen, 1995). It occurs about 100 to 250 ms after stimulus onset and is typically reflected by fronto-central negativity and lateral/mastoid positivity (see Näätänen et al., 2007 for a review). It is said to reflect the ability to discriminate sounds (Bishop, 2007) thus allowing to investigate how well different groups of people perceive speech sound differences (Kujala et al., 2007).
Studies investigating speech sound processing show that the MMN is influenced by language experience. Results point towards the main finding that the MMN is stronger for familiar compared to unfamiliar speech stimuli (Näätänen et al., 1997; Winkler et al., 1999; Cheour et al., 2002; Nenonen et al., 2003; Shestakova et al., 2003; Nenonen et al., 2005), suggesting that our neural networks are shaped by experience (Diaz et al., 2008; Kuhl et al., 2008). However, these studies were looking at language exposure in natural and immersive environments (Näätänen et al., 1997; Winkler et al., 1999; Cheour et al., 2002; Nenonen et al., 2003; Shestakova et al., 2003; Nenonen et al., 2005). As such, it remains unclear whether classroom-based foreign language learning is intensive enough in order to render changes in speech sound processing on the neural level.

1.3 Methodological aspects of EEG/ERP processing

The development of highly sophisticated neuroimaging techniques allows investigating brain-behavior relationships. A widespread and completely non-invasive functional imaging method is the EEG. It is used to investigate spatial and temporal dynamics of neural activity during a wide range of mental states and processes (e.g., Luck, 2005; Michel et al., 2009), in both experimental as well as clinical settings. The ERP reflects the electrophysiological response to a particular event or stimulus.

1.3.1 Topographic ERP approach

Traditionally, ERPs are illustrated as waveforms at certain electrode positions and analyzed based upon their morphology and/or frequency. Waveform analyses typically involve peak amplitude and peak latency analyses, either at single electrodes or electrode clusters. However, modern EEG systems with a large number of electrodes and a high sampling rate also provide additional topographic information. As such, the analyses applied in this thesis were deepened using topographic EEG analyses such as topographic analysis of variance (TANOVA), global field power (GFP) and 3D-centroid analyses. Applying a TANOVA on non-normalized (raw) maps allows detecting differences between two conditions without preselecting a subset of
electrodes (Maurer et al., 2003b; 2008; 2010; Murray et al., 2004; Schulz et al., 2008). This is especially useful in cases where there is only little prior knowledge about the time of occurrence of certain effects. Resulting time windows can reflect either differences in GFP or differences in topography (e.g., Koenig et al., 2011). GFP reflects the ERP map strength and equals the root mean square across all recording electrodes. Moreover, it is independent of the choice of recording reference (Lehmann and Skrandies, 1980). To characterize ERP topography, centroid measures of the positive and negative fields on the scalp surface can be used. These centers of gravity, reflecting the positive and negative fields on the scalp surface (Brandeis et al., 1994), are computed as voltage-weighted locations of all electrodes showing positive or negative values. The positions of the centroids are defined in Talairach space by x-, y- and z-coordinates, which allows analyses in three spatial dimensions “left-right”, “posterior-anterior” and “inferior-superior” (Talairach and Tournoux, 1988, see also Maurer et al., 2010).

The distinction between differences in GFP and differences in topography is important. In the absence of topographic differences, differences in GFP reflect differences in strength of activity, implying that the amount of activation of a neural population engaged in processing the conditions of interest is different. In contrast, differences in topography suggest different configurations of neural sources being activated by the presented stimuli (Michel et al., 2004).

1.3.2 Effects of filtering

Another important methodological aspect relates to filtering of the EEG-data. Filters can be applied to the EEG during data acquisition (analog filtering) or offline (digital filtering), i.e. before or after the averaging process (e.g., Luck, 2005). All of the many different filtering techniques involve removing a portion of the recorded signal, which is very useful for removing activity that is considered to be noise, or to suppress certain signal components in order to focus on others (e.g. Handy, 2005). However, filtering can also change ERPs (e.g., Luck, 2005; Kappenman and Luck, 2010). Therefore, it is important to be aware of the distortions they can introduce to the data. Generally, filters have the ability to suppress or pass certain frequencies. As such, a low-pass filter attenuates high frequencies and passes low frequencies, while a high-pass filter
attenuates low frequencies and passes high frequencies (Handy, 2005; Luck, 2005). The bandpass filter attenuates both high and low frequencies, and passes frequencies within a single range, thus creating a combination of a high-pass and a low-pass filter (Handy, 2005; Luck, 2005). Digital filtering is somewhat arbitrary – thus making it difficult to make the right choice in deciding for the best suited filter-settings with respect to the data and research-question at hand. When applying a typical cognitive experimental paradigm, most of the relevant portions of the ERP waveforms seem to consist of frequencies between 0.01 Hz and 30 Hz (Luck, 2005). As a consequence, in the field of cognitive neuroscience, applying a high-pass filter of 0.01 Hz or 0.03 Hz is a common practice. However, not only does filtering seem to come along with certain risks (VanRullen, 2011; Rousselet, 2012), but depending on the filter settings used, some components of interest could be influenced in a way that certain effects can – or cannot, be visualized in the ERP. A good example for such an effect comes from a sleep study investigating the impact of different filter settings on the MMN (Sabri and Campbell, 2002). It could be shown that increasing the high pass filter to 3 Hz leads to a visualization of the MMN by attenuating the slow-wave positivity. This is methodologically very interesting in light of several studies, which have shown a positive mismatch response in children instead of a typical MMN.

1.4 Conclusion

Audiovisual integration plays a crucial role in learning to read and has thus been investigated in several studies, which have shed light on basic processes of print - speech sound associations. However, the majority of these studies investigated audiovisual integration in adults (e.g., Raij et al., 2000; van Atteveldt et al., 2004; 2007a), and importantly, they used single letters and phonemes instead of entire words (Blau et al., 2010; Froyen et al., 2008; Mittag et al., 2013; van Atteveldt et al., 2004; 2007a). Taking into account that linking print to speech is especially important in reading acquisition, it is interesting to investigate audiovisual integration in children who are at the early stages of learning to read. To this end, we investigated audiovisual
integration of entire (familiar and unfamiliar) words in children who had only one year of reading experience. Given that little knowledge is available yet on the integration of entire words on the neural level, a data driven topographic EEG approach (TANOVA, GFP, 3D-centroids) was applied, not only taking advantage of the high temporal resolution but also of the additional topographic information that EEG systems with high-density electrode net/caps provide.

Research on phonological processing in the context of language acquisition has shown that experience shapes our neural networks (e.g. Kuhl et al., 2008). In order to investigate the neural mechanisms underlying speech sound processing of native and non-native speech sounds, an MMN paradigm was used. The main question was whether classroom-based language learning leads to changes in the MMN towards the non-native speech sound, as was previously shown in studies investigating foreign language processing in an immersive school setting (e.g. Winkler et al., 1999). To this end, longitudinal behavioral and electrophysiological data was acquired from children shortly before they started to learn English as a foreign language at school, and again one year later. Based on studies showing a positive mismatch response in children instead of a typical MMN, different filter analyses were applied resulting in methodologically interesting findings clarifying some inconsistencies in the field of developmental MMN research.

This research is directed towards delivering additional knowledge about basic processes involved in language that in turn can give rise to practical benefits for society in the future, but also towards seizing methodologically interesting topics in the area of EEG/ERP analyses.
2. Integration of Spoken and Written Words in Beginning Readers: A Topographic ERP Study

Lea B. Jost¹,², Aleksandra K. Eberhard-Moscicka¹,², Christine Frisch¹, Volker Dellwo³, Urs Maurer*¹,²

¹Department of Psychology, Division of Cognitive Neuroscience, University of Zurich, Switzerland
²Neuroscience Center Zurich, University of Zurich and ETH Zurich, Switzerland
³Phonetics Laboratory, University of Zurich, Switzerland

3. Native and non-native speech sound processing and the neural mismatch responses: a longitudinal study on classroom-based foreign language learning

Lea B. Jost¹,², Aleksandra K. Eberhard-Moscicka¹,², Georgette Pleisch¹,³, Veronica Heusser¹, Daniel Brandeis³,⁴,⁵, Jason Zevin⁶, Urs Maurer¹,²,⁷

¹ Department of Psychology, Division of Cognitive Neuroscience, University of Zurich, Switzerland
² Neuroscience Center Zurich, University of Zurich and ETH Zurich, Switzerland
³ Department of Child and Adolescent Psychiatry, University of Zurich, Switzerland
⁴ Zurich Center for Integrative Human Physiology (ZIHP), Zurich, Switzerland
⁵ Department of Child and Adolescent Psychiatry and Psychotherapy, Central Institute of Mental Health, Mannheim, Germany
⁶ Departments of Psychology and Linguistics, University of Southern California, USA
⁷ Department of Psychology, The Chinese University of Hong Kong, China

Manuscript published in Neuropsychologia, 2015, 72, 94-104.
4. General discussion

This thesis sheds light on aspects of audiovisual integration and phonological processing. Moreover, it provides insights into methodological aspects of EEG/ERP analyses. In the following sections, an overview of the main results and conclusions will be provided and a short outlook on future research will be given.

4.1 Audiovisual integration

Results from study 1 show that audiovisual integration takes place in children after only one year of formal reading instruction at school. These observations of multisensory integration are supported by the behavioral results, which showed that children detected targets more accurately and faster when they were presented simultaneously in both modalities (AV) compared to when they were presented only visually (V) or only auditory (A). Such multisensory facilitation (Molholm et al., 2004) has also been found in other studies investigating audiovisual integration (e.g., Besle et al., 2004; Kronschnabel et al., 2014; Raij et al., 2000) and interpreted along the “co-activation model” (Miller, 1986). According to this model, the unimodal speech inputs interact, thus facilitating a faster response (see also Besle et al., 2004; Raij et al., 2000). As such, the audiovisual facilitation effect is sought to reflect a combined and interacting process of the unimodal input, rather than a separate processing, where the faster of the two processes to finish triggers a motor response, as suggested by the “race model” (Raab, 1962).

Even though several studies showed associations between reading skills and audiovisual integration (Blau et al., 2010; Froyen et al., 2011), no robust relations were found in study 1. Certain methodological issues could account for the lack of such an association, as for example aspects related to the ERP analyses (i.e. using global map descriptions) or to the experiment (e.g. use of a reading task that is not explicit enough). However, it is also possible, that after only one year of reading training the reading mechanisms are not yet automatized. It was suggested previously, that it takes several years of reading instruction and practice before first signs of automatic integration of letters and speech sounds can be found in normally developing
children (Blomert and Froyen, 2010). Taking into account that we used entire words characterized by a larger lexical complexity than letters and speech sounds, a lack of automation seems plausible.

Audiovisual stimuli (AV) resulted in a weaker neural response compared to summated unimodal responses (SumA+V), independent of the familiarity of the word presented, i.e. whether the stimuli were German words or unfamiliar English words. In contrast to this suppression effect (SumA+V > AVM), congruency effects (AVN ≠ AVM) were only found for German, but not for unfamiliar English words. The early time window reflecting congruency effects for familiar, but not for unfamiliar words (160-204 ms) extends previous findings by showing that lexical-semantic information not only influences early visual word processing (Hauk and Pulvermuller 2004; Sereno et al. 1998; Skrandies 1998; Wirth et al. 2007), but also audiovisual processing. The time of occurrence and the topography of the second congruency effect (500-600 ms) are in line with previously reported N400 effects (Schulz et al., 2008; 2009; Kutas and Federmeier, 2011 for a review), suggesting that matching processes during audiovisual integration are similar to processes involved in semantic priming. The absence of this effect for unfamiliar words suggests that audiovisual integration was restricted to stimuli with semantic content. Also the later congruency effects (1,032–1,108 ms; 1,164–1,188 ms) were modulated by semantic information. Similar results have been reported in semantic priming studies where a late positivity was associated with semantic violations, reflecting integration difficulties and associated reprocessing costs (e.g., Van Petten and Luka, 2012). It thus seems that increased reprocessing of non-matching audiovisual stimuli is only elicited for meaningful stimuli. Hence, the congruency effects for familiar but not for unfamiliar words can be attributed to the lack of meaning of the unfamiliar English words, presumably related to lexical-semantic processing.

4.2 Phonological processing

Based on findings showing that neural networks are shaped by our experience (e.g., Díaz et al., 2008, Kuhl et al., 2008), study 2 investigated the influence of classroom-based foreign language learning on processing native and non-native speech sounds indicated by the MMN.
Specifically, it was studied whether one year of classroom-based foreign language instruction is sufficient in order to find evidence of language-specific memory traces in the cortex (Näätänen et al., 1997, Cheour et al., 2002). In line with the neural commitment theory (Kuhl, 2004; Kuhl et al., 2008), children’s MMN was stronger for the native than the non-native speech sound contrast. However, the same pattern was found after one year of language learning, indicating that English instruction was not intensive enough in order to render changes on the neural level. Importantly, even though the stronger activation for the native compared to the non-native speech sound contrast only reached trend level, the interaction effect with the factor anterior-posterior was significant. Moreover, also at lateral electrodes, there was significantly stronger activation for the native compared to the non-native speech contrast. Furthermore, the TANOVAs revealed overall stronger activation for the native, compared to the non-native speech contrast (significant in 3rd grade, trend in 1st grade). To conclude, there is evidence coming from different analyses verifying the result of a stronger MMN for the native than for the non-native speech contrast.

A possible explanation for the result showing that the MMN towards the non-native speech contrast did not change despite one year of classroom-based foreign language learning, can be found in the neural commitment theory, which states that children’s ability to discriminate speech sounds of foreign languages decreases with age when acoustic prototypes for native phonemic categories are being established (Kuhl, 2004; Kuhl et al., 2008). Neural commitment to the native language could thus be interfering with the processing of the foreign language (e.g. Zhang et al., 2005). The children in study 2 had only two lectures of English per week and were taught by teachers, whose mother tongue was not English. It is thus possible that immersive learning and a more natural setting are a prerequisite to provide qualitatively and quantitatively sufficient input in order to render changes on the neural level.

As most of the previous studies investigating the effect of immersion or classroom-based language learning are from Finnish labs (Shestakova et al., 2003; Peltola et al., 2003; Nenonen et al., 2003; Nenonen et al., 2005), looking at children with a different language background (in this case German), complements current knowledge. Moreover, in the light of current political
and pedagogical debate about foreign language exposure at the primary school level, the topic of early foreign language learning is of particular relevance. Furthermore, longitudinal studies on children learning a foreign language at a regular school where no immersion learning is practiced are very rare. However the objective estimation of the teaching effectiveness with the help of neural correlates could potentially be very useful.

In both study 1 and study 2, a “null-result” (in addition to other significant results) was found for effects that could have been expected based on previous studies. In study 1, the expected correlation between reading fluency and audiovisual integration was absent, and in study 2 the MMN towards the non-native speech contrast did not change despite one year of classroom-based foreign language instruction. As has been documented by Thornton & Lee (2000), there is a risk for a publication bias related to null results, which can for example arise from the researcher himself, who decides not to submit the results because they do not appear to be interesting enough, or which can also arise from the tendency of journals to reject negative results. Nevertheless, the “null-effects” found in study 1 and study 2 contain important information. Moreover, the application of methodologically interesting analyses (TANOVA, GFP, 3-D centroids, additional filter analyses) revealed other valuable findings. In study 1, the use of audiovisual integration paradigms, which have gained substantial interest in research on reading and dyslexia, offered a way to investigate how the brain relates written language to spoken language. Moreover, applying data-driven topographic EEG analyses not only revealed time windows indicating audiovisual integration, but also allowed to characterize these effects. In study 2, instead of the originally found positive mismatch response in children using a high pass filter of 0.3 Hz, an MMN emerged when applying a high-pass filter of 3 Hz. Thus, there seems to be an overlap of a slow-wave positivity with the MMN, indicating that two concurrent mismatch processes were elicited in children. Adults showed topographic MMN differences between the native and non-native speech sound, indicating that speech sounds in a foreign language are processed by at least partly different neural networks compared to speech sounds of the native language.
4.3 Methodological aspects of EEG/ERP processing

Given that little knowledge is available about how entire audiovisual words are integrated in the brain, we applied data-driven topographic EEG analyses (TANOVA, GFP, and centroid analyses) that are more sensitive than traditional analyses restricted to specific components. Comparisons of topographic ERP maps can be interpreted in terms of intra-cerebral sources either in their strength, location or orientation (Koenig and Gianotti, 2009). In study 1 we identified several time windows that indicated audiovisual integration, and were able to characterize these effects by showing whether they were reflected by differences in global map strength or topography. The use of GFP to quantify the effect of an experimental manipulation has several advantages when dealing with EEG scalp field data. Because all sensors are taken into consideration, the scalp field produced by the source(s) is taken into account to the largest possible extent and false negatives based on partially overlapping scalp fields are unlikely (Koenig et al., 2011). Most importantly, as EEG measurements are based on differences of electric potentials, the signal recorded at a certain electrode is dependent on the choice of reference. However, when GFP is used to quantify a response to a certain stimulus, the measure is reference-independent (Koenig et al., 2011).

In study 2, the results are methodologically significant, as for the first time a new filter analysis was applied on children’s mismatch response data. Using typical filter settings, a positive mismatch response was found in children, as reported in previous studies (Maurer et al., 2003a,b; Ceponiene et al., 2004; Ruhnau et al., 2010; Ruhnau et al., 2013). When applying an additional high-pass filter, a negativity emerged with typical temporal and topographic MMN characteristics. Given that the MMN in adults did not change markedly with the 3 Hz high-pass filter, it can be assumed that both the children’s MMN and the adults MMN occurred in a similar frequency range. Thus, there seems to be an overlap of the slow-wave positivity with the MMN suggesting that two simultaneous mismatch processes were evoked. By showing that the positive MMR is not simply an MMN with inverted polarity (Maurer et al., 2003b), but more likely...
a distinct process, this result obtained clarifies certain inconsistencies in the field of developmental MMN research.

4.4 Limitations and outlook

The findings presented here provide valuable insights into basic processes of language processing, but also into methodologically important aspects of EEG/ERP analyses. However, study 1 and study 2 are not without limitations and several questions remain open to further investigation. In the following, implications and topics that could be of interest for future research will be discussed.

Study 1 reports data on audiovisual integration of familiar and unfamiliar words in children after one year of reading training. Given the novelty of the paradigm that uses entire words as well as the sample of children at an early stage of reading acquisition, this study provides important insights into the nature of audiovisual integration at the beginning of learning to read. Based on the findings of study 1, where no robust associations between reading fluency and audiovisual integration were found, and given the theoretical importance of grapheme-phoneme conversion for learning to read (Blomert, 2011; Wimmer and Schurz, 2010), it would be interesting to investigate audiovisual integration of entire words in older children who are more advanced readers. This would allow drawing further conclusions whether an association between reading fluency and audiovisual integration is indeed due to lack of automation of the reading process, as suggested in study 1, or whether other factors related to the experiment (e.g. use of a reading task that is not explicit enough) or to the ERP analyses (i.e. using global map descriptions) influenced the results, too.

Moreover, based on studies showing an association between reading difficulties and either phonological processing (e.g., Alonso-Bua et al., 2006; Schulte-Koerne et al., 2001; Kujala et al., 2000; Cheour et al., 2000) or audiovisual integration (Blau et al., 2009; Froyen et al., 2011), it might furthermore be interesting to see whether children showing lower audiovisual integration also show weaker MMNs for native speech sound contrasts. This question could be addressed
with the data from study 1 and study 2, and analyses are planned to be conducted in order to pursue this question.

Furthermore, as mentioned in the methods section of study 1, two kinds of English words were presented; simple English words, which have similar grapheme-phoneme correspondence rules as found in the German language (e.g. “pen”), and complex English words with different grapheme-phoneme correspondence rules compared to the German language (e.g. “eye”). Preliminary data shows that the children in study 1 did not learn all of the English words presented. Thus, unknown simple and complex English words would need to be excluded from further analyses. Hence, there wouldn’t be enough data to investigate German words, simple English words and complex English words separately in order to draw meaningful comparisons. However, it would be very interesting for future studies to investigate to what extent audiovisual integration is associated with the orthographic depth of words. Orthographic depth, which refers to the complexity, consistency or transparency of grapheme-phoneme correspondences of a language (Frost et al., 1987), has been shown to play a critical role in learning to read (e.g. Seymour et al., 2003; Landerl et al., 2013). Comparing audiovisual integration of words with a deep (i.e. inconsistent or opaque) orthography to audiovisual integration of words with a shallow (i.e. consistent or transparent) orthography, would indicate how easy or difficult it is for children to translate letter strings into a phonological code depending on the orthographic depth of the words (see also Richlan, 2014).

Study 2 reports longitudinal data of children shortly before they started to learn English in primary school and of the same children one year later. The main result is that children’s MMN was larger for native than non-native speech sound deviance in agreement with the neural commitment theory (Kuhl, 2004), but also that this did not change after one year of language learning. Thus, the limited intensity of one year of classroom-based foreign language learning did not lead to changes in processing non-native speech sounds, as had previously been demonstrated for immersive learning (e.g. Winkler et al., 1999). Based on the conclusion that the two lessons of English per week are not intensive enough in order to render changes on the neural level, it would be interesting to investigate whether neural changes in processing non-
native speech sounds can be found in older children (e.g. children in 5th or 6th grade) after four or five years of classroom-based English instruction. Moreover, it would be informative to compare the MMN in response to native and non-native speech sound contrasts of children learning a new language in a classroom setting, to the MMN of children learning a new language in an immersive setting. As other studies investigating the effect of immersion-education on phonological processing have shown learning-related changes in the MMN, it would be interesting to see whether the lack of a similar effect in study 2 is indeed due to low intensity. A direct comparison between children learning English in a classroom setting and children learning in an immersion setting would be a large step towards investigating how much foreign language exposure is needed in order to render changes on the neural level and whether immersion education indeed supports changes on the neural level already at early stages in primary school (i.e. after only one year of English learning).

The standards (/da/) in study 2, which were included in the ERP computation, were not always followed by another standard. As a result, the mismatch response that occurs after 450 ms overlaps with the next stimulus and could thus be influenced by this difference. As the main results of study 2 are the MMN differences between native and non-native speech contrasts, as well as the effect of the filtering on the early positive mismatch response in children, the analyses were restricted to the time range up to 450 ms. However, it would be interesting to see whether there are effects of learning a foreign language on later ERP components than the typical MMN. It has previously been proposed that the LDN (late deviant negativity) potentially plays an important role in the process of language learning and more generally, audition-based learning processes (Zachau et al., 2005, Jakoby et al., 2011). Based on the assumption that the LDN can serve as an index of formation of a phonological representation (Barry et al., 2009), it would be informative to test whether LDN changes as a consequence of acquiring a new language. This would either require a new analysis of the ERP data from study 2, or a longer SOA in future experiments.
In terms of methodological aspects, the filter analysis was able to resolve some inconsistencies in the developmental MMN literature. However, filters should be applied cautiously and future research shall address the issue of how filter effects influence the ERP data.

Altogether, the present thesis sheds light onto audiovisual integration at early stages of learning to read and phonological processing in the context of classroom-based foreign language learning. The aim of investigating learning a foreign language in school longitudinally is an ambitious undertaking and study 2 was one of the first studies on classroom-based foreign language learning. Moreover, by addressing several methodological aspects of EEG/ERP analyses, this thesis also offers insights for future research on topographical issues of ERP analyses and issues related to effects of different filter-settings.
References


Alho K (1995) Cerebral generators of mismatch negativity (MMN) and its magnetic counterpart (MMNm) elicited by sound changes. *Ear and Hearing* 16, 38-51


Bishop DVM (2007) Using Mismatch Negativity to Study Central Auditory Processing in Developmental Language and Literacy Impairments: Where Are We, and Where Should We Be Going? *Psychological Bulletin* 133, 4, 651-672


Lyster R & Genesee F (2012) Immersion Education. The Encyclopedia of Applied Linguistics


Miller J (1986) Timecourse of coactivation in bimodal divided attention *Perception & Psychophysics 40*, 331–343


Nenonen S, Shestakova A, Huotilainen M, Näätänen R (2005) Speech-sound duration processing in a second language is specific to phonetic categories. *Brain and Language* 92, 26-32

Nikolov M (Ed.) (2009) *The age factor and early language learning (40).* Walter de Gruyter


Torgesen JK, Wagner RK, Rashotte CA (1999) TOWRE - Test of word reading efficiency. Austin, Texas: PRO-ED


Acknowledgements

First of all I would like to thank my direct supervisor Prof. Dr. Urs Maurer who introduced me to the world of EEG and language research. His patience in teaching, his profound knowledge and his openness to new ideas will always be exemplary for me. I appreciate his support and I am very grateful for having gotten the chance to be part of his project.

I’m also most grateful to Prof. Dr. Volker Dellwo and to Dr. sc. nat. Silvia Brem for their supervision and for their helpful and inspiring comments. As members of my PhD steering committee they were a very valuable support and I’ve always experienced meetings as enriching. I am also much obliged to Prof. Dr. Klaus Oberauer for his interest in my research, for agreeing to be the independent expert of my PhD thesis and for his valuable inputs on the MMN-paper.

My gratitude also goes to Margit Raith, who was not only indispensable for matters concerning the project, but who also always had an open ear for us.

Special thanks go to the Master students Georgette Pleisch, Lynn Fehlbaum, Veronica Heusser, Barbara Rindlisbacher, Désirée Eberle, and Christine Frisch who wrote their master thesis within the project and who helped with data collection, data preprocessing and data analyses. All of them did a fantastic job and I enjoyed very much working with them.

Special thanks also to all the interns, not only for their help with data collection, but also for their enthusiasm and for the creation of a good and often inspiring atmosphere.

I am also most grateful to Simone Pfenninger who helped developing the English tests. I very much appreciated working with her.

Moreover, I would also like to thank the Swiss National Science Foundation for the financial support. Special thanks also to the schools who agreed for doing recordings in extra rooms provided by them.

I am also most grateful to all the children, parents and adults who participated in the study and who made this work possible.

Last but not least, I would like to thank Aleksandra Eberhard-Moscicka. I will always remember our “first steps” with the EEG, the periods when we were studying for exams, our unbeatable lunch breaks, the superb time we had in Geneva...Thank you for being not only a great colleague, but also a reliable and trustworthy friend.
Curriculum Vitae

Lea Jost
*16. Mai 1984, Bern

2003

High school graduation
Abitur, Deutsche Schule Kapstadt – South Africa

Academic studies
2007
Bachelor of Psychology, University of Bern
Minor: Criminal Law and Criminology

2010
Master of Science in Psychology, University of Bern
Major: Neuropsychology
Diploma thesis:
Training des Arbeitsgedächtnisses bei Kindern mit
Aufmerksamkeits- und Gedächtnisproblemen
Minor: Social Psychology

Internships
2008/2009
Zentrum für Entwicklungsförderung und pädiatrische
Neurorehabilitation der Stiftung
Wildermeth Biel (Z.E.N.)

01/2010 - 04/2010
Rehabilitation Center Leukerbad,
Division Neuropsychology

05/2010 - 10/2010
Gerontopsychiatrisches Zentrum Hegibach, Zurich
Division Neuropsychology

since 12/2010

Doctoral position, University of Zurich

Affiliation: Cognitive Neuroscience
Advisors: Prof. Dr. U. Maurer
Prof. Dr. K. Oberauer
Prof. Dr. V. Dellwo
Dr. sc. Nat. S. Brem
PhD programs
- PhD Program, Psychology, University of Zurich
- ZNZ International PhD Program in Neuroscience, Swiss Federal Institute of Technology ETH & UZH

Teaching
Practical part of the seminar “Topographic analyses of EEG/ERP data”, Department of Psychology, UZH

Supervision of Master students who completed their thesis within our project

Conference poster presentations
- Deutsches EEG/EP Mapping Meeting (DMM), Giessen
- Cognitive Neuroscience Society Meeting (CNS), San Francisco
- International Conference on Basic and Multimodal Clinical Imaging (BaCi), Geneva
- Congress of the Swiss Psychological Society (SGP), Basel
- Donders Discussions, Nijmegen

Title PhD Thesis
Audiovisual integration of words and processing of native and non-native speech sounds

Invited talk
Reading in the Brain – Guest Lecture in the Course ‘Second Language Learning and Teaching’ at the English Department of the UZH

Publications

