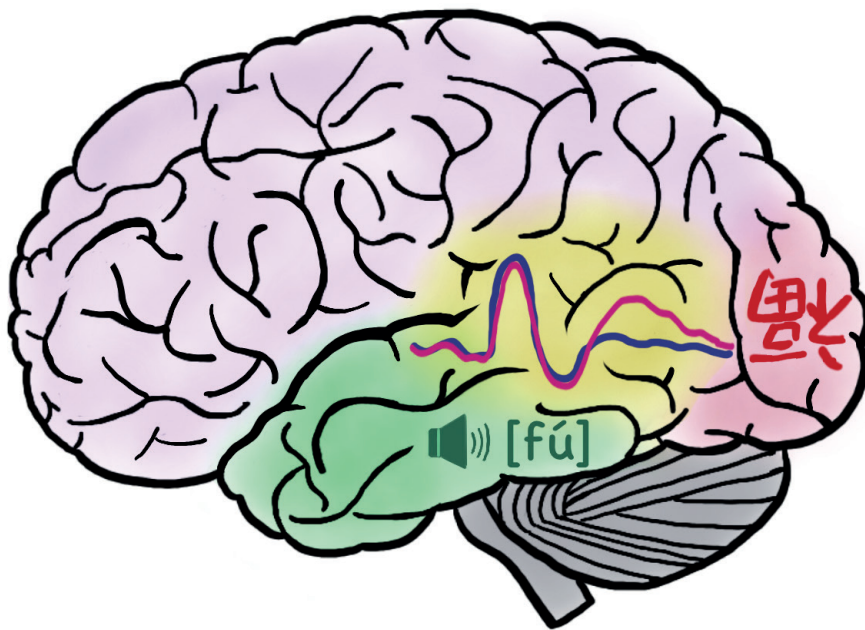


JYU DISSERTATIONS 249

Weiyong Xu

Brain Activity Changes Related to Learning of Audiovisual Associations in Reading



UNIVERSITY OF JYVÄSKYLÄ
FACULTY OF EDUCATION AND
PSYCHOLOGY

JYU DISSERTATIONS 249

Weiyong Xu

**Brain Activity Changes Related
to Learning of Audiovisual
Associations in Reading**

Esitetään Jyväskylän yliopiston kasvatustieteiden ja psykologian tiedekunnan suostumuksella
julkisesti tarkastettavaksi elokuun 14. päivänä 2020 kello 12.

Academic dissertation to be publicly discussed, by permission of
the Faculty of Education and Psychology of the University of Jyväskylä,
on August 14, 2020 at 12 o'clock noon.



JYVÄSKYLÄN YLIOPISTO
UNIVERSITY OF JYVÄSKYLÄ

JYVÄSKYLÄ 2020

Editors

Noona Kiuru

Department of Psychology, University of Jyväskylä

Ville Korkiakangas

Open Science Centre, University of Jyväskylä

Cover picture by Xueqiao Li.

Copyright © 2020, by University of Jyväskylä

Permanent link to this publication: <http://urn.fi/URN:ISBN:978-951-39-8220-1>

ISBN 978-951-39-8220-1 (PDF)

URN:ISBN:978-951-39-8220-1

ISSN 2489-9003

ABSTRACT

Xu, Weiyong

Brain activity changes related to learning of audiovisual associations in reading
Jyväskylä: University of Jyväskylä, 2020, 70 p.

(JYU Dissertations

ISSN 2489-9003; 249)

ISBN 978-951-39-8220-1 (PDF)

Learning to connect letters or characters of written scripts to their corresponding sounds is crucial for reading acquisition. In alphabetic languages, the letter-speech sound integration has been shown to have a protracted developmental trajectory, and failure to reach an automatic level of audiovisual integration was correlated with reading difficulties. This dissertation aims to systematically investigate the audiovisual integration process in learning to read using magnetoencephalography, by extending the previous findings on alphabetic language to logographic language, and furthermore, by examining the learning of grapheme-phoneme association during initial learning stages. Study I aimed to investigate the audiovisual integration process in a logographic language (Chinese). This audiovisual integration involved the left superior temporal cortex in Chinese, which is similar to findings in alphabetic languages. In addition, it also activated the left inferior frontal regions, which are related to the processing of additional semantic information embedded in Chinese characters. Study II correlated various brain indices of audiovisual processing with reading-related cognitive measures in children at varying stages of reading. It demonstrated that the auditory late component is closely related to rapid automatized naming and phonological processing skills. Moreover, the multisensory interaction effect was observed mainly in temporoparietal regions, and brain responses in some of these regions were further associated with children's reading and writing abilities. Study III simulated the initial learning of grapheme-phoneme associations in adults. The results from Study III highlighted the dynamic characteristics of audiovisual learning and provided a more refined model of grapheme-phoneme learning in reading acquisition. Overall, the findings from this dissertation showed evidence that audiovisual processing is dynamic during initial learning and memory consolidation of cross-modal associations. Furthermore, audiovisual processing is less automatic in children and is linked to their reading-related cognitive skills. Finally, there are some universal audiovisual processing brain regions and mechanisms across languages that are complemented by additional regions related to processes of distinct linguistic features in different types of scripts.

Keywords: audiovisual integration, language learning, child brain, magnetoencephalography, reading

TIIVISTELMÄ (FINNISH ABSTRACT)

Xu, Weiyong

Audiovisuaalisten yhteyksien oppimiseen liittyvät aivot toiminnan muutokset lukemaan oppimisen yhteydessä

Jyväskylä: University of Jyväskylä, 2020, 70 p.

(JYU Dissertations

ISSN 2489-9003; 249)

ISBN 978-951-39-8220-1 (PDF)

Kirjainten tai merkkien yhdistäminen niitä vastaaviin äänteisiin on ratkaisevan tärkeää lukemaan oppimiselle. Aakkosia käyttävissä kielissä kirjainten ja äänteiden yhdistämisellä on havaittu hidaskas kehityskaarit ja tämän on osoitettu korreloivan lukemisvaikeuksien kanssa. Tässä väitöstutkimuksessa lukemaan oppimisen audiovisuaalista integraatioprosessia tutkitaan magnetoenkefalografian (MEG) avulla. Aakkosia käyttävissä kielissä saadut tulokset laajennetaan koskemaan logografista eli sanakirjoitusta käyttävää kieltä. Lisäksi tutkitaan grafeemi-foneemi-vastaavuuden omaksumista lukemaan opettelu-alkuvaiheissa. Tutkimus 1 kartoittaa audiovisuaalista integraatioprosessia logografisessa kielessä (kiina). Tulosten perusteella kiinan kielen merkkejä opeteltaessa audiovisuaalinen integraatio aktivoi vasemman ylemmän ohimolohkon alueita, mikä vastaa aakkosia käyttävien kielten tuloksia. Lisäksi se aktivoi vasemmanpuoleiset alemmat otsalohkoalueet, jotka ovat yhteydessä kiinan merkkien sisältämään semanttiseen lisäinformaatioon. Tutkimuksessa 2 suoritettiin korrelaatioanalyysi erilaisten audiovisuaaliseen prosessointiin liittyvien indeksien ja lukemiseen liittyvien kognitiivisten testitulosten välillä lapsille, jotka olivat eri vaiheissa lukemaan oppimisessa. Tämä osoitti, että myöhäinen auditiivinen heräteaste on yhteydessä nopeaan automaattiseen nimeämiseen ja fonologisiin prosessointitaitoihin. Lisäksi moniaistinen vuorovaikutusefekti ilmeni pääasiassa temporoparietaalisilla alueilla, ja aivojen heräteasteet osasta näistä alueista olivat edelleen yhteydessä lasten luku- ja kirjoitustaitoon. Tutkimuksessa 3 tutkittiin grafeemi-foneemi-vastaavuuksien varhaista oppimista opettamalla uusia yhteyksiä aikuisille. Kolmannen osakokeen tulokset toivat esiin aistien väliseen oppimisen dynaamiset piirteet tarjoten kehittyneemmän grafeemi-foneemi-vastaavuuksien oppimisen mallin. Kaiken kaikkiaan tutkimustulokset osoittavat, että audiovisuaalinen prosessointi on dynaamista aistien välisen yhteyksien varhaisessa oppimisessa ja niiden muistamisen vahvistumisessa. Lisäksi audiovisuaalinen prosessointi on lapsilla vähemmän automaattista ja yhteydessä heidän lukemiseen liittyviin kognitiivisiin taitoihinsa. On myös olemassa universaaleja, kaikille kielille yhteisiä audiovisuaalisen prosessoinnin aivoalueita ja -mekanismeja. Niitä täydentävät alueet, jotka liittyvät tiettyjen kielellisten piirteiden prosesseihin erilaisissa kirjoitusjärjestelmissä.

Asiasanat: audiovisuaalinen integraatio, kielen oppiminen, lapsen aivot, magnetoenkefalografia, MEG, lukeminen

Author

Weiyong Xu
Department of Psychology
University of Jyväskylä
P.O. Box 35
40100 University of Jyväskylä, Finland
weiyong.w.xu@jyu.fi
<https://orcid.org/0000-0003-4453-9836>

Supervisors

Professor Jarmo A. Hämäläinen
Department of Psychology
University of Jyväskylä, Finland

Professor Paavo H. T. Leppänen
Department of Psychology
University of Jyväskylä, Finland

Professor Robert Oostenveld
Donders Institute for Brain, Cognition and Behaviour
Radboud University, Netherlands
NatMEG, Department of Clinical Neuroscience
Karolinska Institutet, Sweden

Reviewers

Professor Milene Bonte
Department of Cognitive Neuroscience
Maastricht University, Netherlands

Professor Urs Maurer
Department of Psychology
The Chinese University of Hong Kong, Hong Kong

Opponent

Professor Milene Bonte
Department of Cognitive Neuroscience
Maastricht University, Netherlands

ACKNOWLEDGEMENTS

First of all, I would like to express my deepest gratitude to my supervisor Professor Jarmo Hämäläinen for his dedicated support and guidance throughout my Ph.D. studies. Your patience and encouragement have given me confidence in this new and exciting research field. I am also grateful to my second supervisor Professor Paavo Leppänen for being an excellent coordinator in the ChildBrain project and for his insightful advice. I would like to thank my third supervisor Professor Robert Oostenveld for his crucial support in the MEG data analysis and statistical methods. I wish to thank my opponent Professor Milene Bonte and the external reviewer Professor Urs Maurer for their valuable time and effort in reviewing the dissertation. I must also thank all the participants in the three studies and research assistants who helped me in data collection. Furthermore, I had great pleasure working with all the co-authors of the papers included in this dissertation.

I want to thank all my colleagues at the Department of Psychology, University of Jyväskylä, for giving me peer support and filling the working time with so much fun and inspiration. Special thanks to our ChildAble team: Praghajieeth Santhana Gopalan, Sam van Bijnen, Orsolya Kolozsvari, Natalia Louleli, and Najla Azaiez Zammit Chatti. It has been my privilege to know you and spend so much memorable time together with you. I would also like to thank all my colleagues (Praghajieeth Santhana Gopalan, Caroline Beelen, Raul Granados, Gloria Romagnoli, Cecilia Mazzetti, Diandra Brkić, Sam Van Bijnen, Anna Samsel, Abinash Pant, Vân Phan, Simon Homölle, Maria Carla Piastra, Marios Antonakakis, and Amit Jaiswal) and PIs in the ChildBrain consortium for all for the wonderful scientific and social experience we shared.

I am grateful to Lin Xie, Chao Wang, Juho Strömmer, Professor Jin Xu, and Professor Fengyu Cong for their kind help and encouragement, which motivated me to pursue a Ph.D. in Finland. Thanks should also go to all my Chinese friends (Xueqiao Li, Jia Liu, Yongjie Zhu, Chaoxiong Ye, Qianru Xu, Lili Tian, and many others) in Jyväskylä who have helped me and shared their delicious food with me during these past years. I am also very thankful to Chunhan Chiang for his warm-hearted support and friendship.

Finally, I would like to thank my parents, who have nurtured me and always supported me.

This dissertation, including all three studies, have been supported by the European Union projects ChildBrain (Marie Curie Innovative Training Networks, no. 641652), Predictable (Marie Curie Innovative Training Networks, no. 641858), the Academy of Finland (MultiLeTe #292466) and the Department of Psychology, University of Jyväskylä.

Jyväskylä, 20 February 2020
Weiyong Xu

LIST OF ORIGINAL PUBLICATIONS

- I Xu, W., Kolozsvári, O. B., Oostenveld, R., Leppänen, P. H. T., & Hämäläinen, J. A. (2019). Audiovisual processing of chinese characters elicits suppression and congruency effects in MEG. *Frontiers in human neuroscience*, 13,18.
<https://doi.org/10.3389/fnhum.2019.00018>
- II Xu, W., Kolozsvari, O. B., Monto, S. P., & Hämäläinen, J. A. (2018). Brain responses to letters and speech sounds and their correlations with cognitive skills related to reading in children. *Frontiers in human neuroscience*, 12, 304.
<https://doi.org/10.3389/fnhum.2018.00304>
- III Xu, W., Kolozsvari, O. B., Oostenveld, R., & Hämäläinen, J. A. (2020). Rapid changes in brain activity during learning of grapheme-phoneme associations in adults. *NeuroImage*, 117058.
<https://doi.org/10.1016/j.neuroimage.2020.117058>

Taking into account the instructions given and comments made by the co-authors, the author of this thesis contributed to the original publications as follows: he designed the experiments, collected the MEG data, conducted the analyses and wrote the manuscripts of the three studies.

FIGURE

FIGURE 1	Statistical results of suppression and congruency effects at the sensor and source levels for the Chinese and Finnish groups.	31
FIGURE 2	Brain regions that showed significant suppressive interaction effects ($A + V > AVC$).....	34
FIGURE 3	The spatiotemporal cluster-based statistical results for the Learnable vs. Control comparisons (auditory, visual, audiovisual interaction) and contrasts of the brain responses to different learning cues at different learning stages in two days.	37
FIGURE 4	Schematic diagram of the possible network involved in the learning of letter-speech sound associations.	49

TABLE

TABLE 1	Summary of methods in all three studies.....	28
TABLE 2	Summary of the clusters exceeding the randomization distribution under H_0 for suppression and congruency effects at sensor and source levels in the Chinese ($N = 12$) and the Finnish ($N = 13$) groups.	31
TABLE 3	Significant partial correlations between sensor brain responses and reading-related cognitive skills (controlling for the effect of age)...	32

CONTENTS

ABSTRACT

TIIVISTELMÄ (FINNISH ABSTRACT)

ACKNOWLEDGEMENTS

LIST OF ORIGINAL PUBLICATIONS

FIGURES AND TABLES

CONTENTS

1	INTRODUCTION	11
1.1	Multisensory integration	11
1.2	Auditory and visual sensory pathways	12
1.2.1	The Auditory Pathway	12
1.2.2	The Visual Pathway	13
1.3	Audiovisual integration.....	14
1.3.1	Analysis approach.....	14
1.3.2	Brain regions involved in audiovisual processing.....	15
1.3.3	Timing of brain activity related to audiovisual processing	16
1.3.4	Audiovisual integration in alphabetic languages	17
1.3.5	Behavioral correlates of audiovisual integration.....	18
1.4	Audiovisual learning in the human brain.....	18
1.4.1	Letter-speech sound learning	18
1.4.2	Training studies on learning of artificial grapheme-phoneme associations	19
1.5	Aims of the research.....	20
2	METHODS	23
2.1	Participants	23
2.2	Behavioral measures	23
2.3	Stimuli and task	24
2.4	MEG and MRI data acquisition	25
2.5	Data analysis.....	26
2.6	Statistical analysis	27
3	RESULTS	29
3.1	Study I	29
3.2	Study II.....	31
3.3	Study III.....	34
3.3.1	Congruency effects in the pSTS	35
3.3.2	Cortical responses to unimodal stimuli and audiovisual interaction (Learnable vs. Control).....	35
3.3.3	Cortical responses to different learning cues.....	35
3.3.4	Correlations between cognitive skills and learning speed	38

4	DISCUSSION	39
4.1	Audiovisual integration in logographic languages	40
4.2	Audiovisual processing and its behavioral correlates in children	42
4.3	Audiovisual learning in the human brain.....	44
4.4	General discussion.....	47
4.5	Limitations	50
4.6	Future directions.....	52
	SUMMARY IN FINNISH	54
	REFERENCES.....	56
	ORIGINAL PAPERS	

1 INTRODUCTION

Learning grapheme–phoneme association is a crucial step in reading acquisition in alphabetic languages, which requires the brain to associate the visual representation of letters to the auditory representation of phonemes. This kind of audiovisual learning relies on the ability to integrate auditory and visual information from the unisensory brain regions and to form a multisensory associative representation. The audiovisual associations are strengthened through repetition and reading practice, becoming faster and more automatic, which enables fluent reading. Research has shown that this process could take years to be fully automated and the failure of which has been related to reading difficulties (dyslexia) in children. The current dissertation aims to systematically investigate the brain mechanisms of audiovisual processing at different stages of reading acquisition, with a particular interest in brain dynamics during initial learning. In addition, previous studies on audiovisual integration have been conducted mostly on alphabetic languages in which the letter–sound mapping is rather simple and consistent. Less is known about the audiovisual processing in other types of scripts. For example, Chinese is a logographic script in which characters map onto mostly syllable-based morphemes in the spoken language. Therefore, compared to alphabetic languages, similar cross-modal associations have to be learned in Chinese but with an additional link to the meanings embedded in the characters. The universal and orthography-dependent brain mechanisms of cross-modal integration are further explored in a logographic script (Chinese) in this dissertation.

1.1 Multisensory integration

In everyday life, we constantly receive and integrate information from multiple sensory modalities. Multisensory integration is essential for high-level human cognition since the ability to form a coherent multimodal representation of objects is necessary for better understanding and interpretation of the external

world. Among all the sensory modalities in humans, auditory and visual pathways convey crucial sensory inputs which contain dense and complex information. For example, our ability to converse relies on the ability to integrate mainly the auditory (sounds) and visual (lip movements) inputs and form a multisensory representation of the speech sound in the brain. Therefore, it is not surprising to find that the auditory and visual inputs are integrated and processed at various levels and stages in the brain.

Visual inputs affect the perception of heard speech. For example, in a face-to-face conversation (lipreading), the speech perception is markedly enhanced if the speaker's lip movements are available and this is particularly important in noisy conditions (Bernstein, Auer, & Takayanagi, 2004; Sumbly & Pollack, 1954), which involves segregation of different streams of sound sources (Bregman, 1994). Another example of this audiovisual interaction is the classic perceptual phenomenon called the McGurk effect (McGurk & MacDonald, 1976), in which, for example, the sound of /pa/ overlaid on a visual /ka/ would induce the fusion perception of /ta/.

Spoken language includes speech perception and production, which are products of biological evolution. However, written scripts are more recent cultural inventions, which have existed for only a few thousand years (Liberman, 1992). In learning to read, speech sounds are associated with arbitrary letters/characters in different kinds of written languages. This implied that the brain has to adapt or change parts of the existing circuits for the challenge of associating novel symbols with speech sounds. Research (Stekelenburg, Keetels, & Vroomen, 2018) has found that compared to visual speech, the visual texts have much weaker effects on the sound processing, which might be due to the fact that the audiovisual associations in written scripts are arbitrary and require explicit learning later in life. A prolonged neurocognitive developmental trajectory of letter-speech sound processing has been observed by mismatch negativity (MMN) studies in children learning to read (Froyen, Bonte, van Atteveldt, & Blomert, 2009). It has been demonstrated that only after obtaining extensive reading experience does the adult brain show signs of automatic letter-speech integration (Froyen, van Atteveldt, Bonte, & Blomert, 2008) and efficient neural associations between the visual text and auditory cortical representation (Bonte, Correia, Keetels, Vroomen, & Formisano, 2017).

1.2 Auditory and visual sensory pathways

1.2.1 The Auditory Pathway

The auditory pathway involves transforming air pressure waves into a neural code of the sound. The sound wave (changes in the air pressure) reaches the outer ear and is funneled through the ear canal to the eardrum. Then the eardrum vibrates the auditory ossicles (delicate bones in the middle ear), which amplify the sound signal and send it to the cochlea in the inner ear. Hair cells within the

cochlea transform the sound vibrations into neural electrical signals which travel through the auditory nerve and subcortical nuclei to the auditory cortex in the brain. The primary auditory cortex is located in the superior temporal gyrus (STG) and expands into the area of the lateral sulcus and the transverse temporal gyrus (also known as Heschl's gyrus). Different types of neurons within the auditory cortex show distinct response properties for encoding intensity, frequency, timing, and spatial information of the sound (Moerel, De Martino, & Formisano, 2014). In addition, the human auditory cortex is adapted to process complex features, such as rhythm, harmony, melody, and timbre in complex sounds, which are common in human speech and music. The perception of the difference or change in the sound stimulus is reflected by the evoked potentials (EP) measured on the scalp surface using the electroencephalogram (EEG) (Eggermont & Ponton, 2002). For example, the auditory mismatch negativity (MMN) response is elicited by discriminable changes in the repetitive aspect of auditory inputs that are stored in the auditory sensory memory (Näätänen, Gaillard, & Mäntysalo, 1978; Näätänen, Tervaniemi, Sussman, Paavilainen, & Winkler, 2001).

Complex auditory input, such as speech, is processed in distributed brain regions and involves multiple processing stages related to different linguistic features. The initial phase of speech processing includes spectrotemporal analysis in bilateral auditory cortices. Then phonological-level processing and representation are suggested to occur in the middle to posterior parts of the bilateral superior temporal sulcus (STS) with weak lateralization towards the left-hemisphere. A dual-route model of cortical organization for speech processing was first suggested by Hickok and Poeppel (Hickok, 2012; Hickok & Poeppel, 2000). The ventral stream, which includes the superior and middle temporal cortex, is linked to mapping speech input onto conceptual and semantic representations for comprehension; the dorsal stream, which consists of the left-dominant posterior planum temporale and frontal region, is responsible for connecting acoustic speech with articulation in speech production. Various event-related potentials (ERP) (e.g., N100, N400, and P600) have been identified to show evoked responses to the presentation of speech stimuli. These language-relevant ERP components are ordered in time and linked to specific processes of linguistic features. For example, the N100 component (negative peak around 100 ms) has been associated with acoustic processes (Obleser, Lahiri, & Eulitz, 2003), and the N400 component (negative peak around 400 ms) was related to lexical-semantic processes (Kutas & Hillyard, 1980).

1.2.2 The Visual Pathway

Visual inputs from the retina are first forwarded to the lateral geniculate nucleus (LGN) located in the thalamus before arriving at the primary visual cortex (V1). Visual inputs flow through a cortical hierarchy in the human brain. Different visual features are processed through this hierarchical structure (V2, V3, V4, and V5) with increasing neural representations. The level of functional specialization also increased with the increasing complexity of neural representation. Two

distinct pathways have also been proposed for the visual pathway (Mishkin & Ungerleider, 1982): the dorsal stream and the ventral stream. First, the dorsal stream (also known as the “where pathway”) connects to the parietal lobe and is involved in the direction of actions and perceiving location of objects. The ventral stream (or the “what pathway”) goes into the temporal cortices and is activated during visual object recognition, identification, and categorization. Notably, the left ventral occipitotemporal (vOT) area is an important brain region adapted for visual letter string/word processing. The left vOT connects the visual word forms to other language areas of the brain, and shows a posterior to anterior gradient (Lerma-Usabiaga, Carreiras, & Paz-Alonso, 2018; Vinckier et al., 2007), with the posterior part involved in the extraction of visual features and is sensitive to smaller grain sizes (e.g., letters) and the anterior part to larger grain sizes, such as words (Dehaene et al., 2010). In reading acquisition, left vOT develops an abstract representation of symbols that are invariant of the case, font, and size. Furthermore, vOT interacts with spoken language systems, for example, the phonological representations in the temporal cortex (Price & Devlin, 2011).

However, research (Dehaene & Cohen, 2011) has shown that purely visual exposure is not sufficient to induce changes in vOT in relation to written stimuli. Changes only started to emerge with top-down control (Song, Hu, Li, Li & Liu, 2010) and attention (Hashimoto & Sakai, 2004; Yoncheva, Blau, Maurer, & McCandliss, 2010) to the interconnection between the visual and auditory inputs, in other words, by learning the letter–speech sound associations.

1.3 Audiovisual integration

1.3.1 Analysis approach

In general, brain mechanisms of audiovisual interactions have been investigated (Murray & Spierer, 2009; Raij, Uutela, & Hari, 2000; van Atteveldt, Formisano, Goebel, & Blomert, 2004) with audiovisual paradigms which normally use four types of stimuli: unisensory auditory (A), unisensory visual (V), audiovisual incongruent (AVI) and audiovisual congruent (AVC). With this kind of audiovisual experiment design, there are two principal analysis approaches that could be served as indices of audiovisual processing.

The first method is derived from the additive model, in which the audiovisual response is measured against the summation of the unisensory auditory responses and visual responses [AV vs. (A + V)]. This approach is suitable for almost any type of multimodal experiment design with random combinations of unisensory stimuli and has been commonly employed in electrophysiological research on multisensory integration (Calvert & Thesen, 2004; Raij et al., 2000; Sperdin, Cappe, Foxe, & Murray, 2009; Stein & Stanford, 2008). In addition, both sub-additive [AV < (A + V)] and supra-additive [AV > (A + V)] effects could be detected, including the modulation of unisensory brain processes in the unisensory cortical regions and novel brain processes specifically

triggered by the bimodal feature of the stimulus with the assumption that minimal common brain activity was presented across different conditions (Besle, Fort, & Giard, 2004). Both sub-additive and supra-additive cross-modal interaction has been reported in the neurons located in the superior temporal sulcus and superior colliculus in animal studies using an electrophysiological approach (Kayser, Petkov, & Logothetis, 2008; Laurienti, Perrault, Stanford, Wallace, & Stein, 2005; Meredith, 2002; Perrault, Vaughan, Stein, & Wallace, 2005; Schroeder & Foxe, 2002; Stein & Stanford, 2008).

Electro-magneto-encephalography (M/EEG) studies on humans have mostly reported suppressive audiovisual effects (Fort, 2002; Foxe et al., 2000; Jost, Eberhard-Moscicka, Frisch, Dellwo, & Maurer, 2014; Lütkenhöner, Lammertmann, Simões, & Hari, 2002; Molholm et al., 2002; Raji et al., 2000; Schröger & Widmann, 1998; Teder-Sälejärvi, McDonald, Di Russo, & Hillyard, 2002). In fMRI studies, interaction (e.g., [AV vs. (A+V)]) (Calvert et al., 1997; Calvert, Hansen, Iversen, & Brammer, 2001) and conjunction analysis (e.g., [(AV > A) \cap (AV > A) \cap A \cap V]) (Beauchamp, Lee, Argall, & Martin, 2004; van Atteveldt et al., 2004) has been used for identifying brain regions involved in cross-modal integration. However, the results of fMRI studies should be interpreted with caution since the specific criteria of audiovisual integration and the analysis strategy was often not consistent across different studies (Calvert, 2001; van Atteveldt et al., 2004).

Another analysis method for investigating audiovisual integration is the congruency comparison (Hein et al., 2007; Jones & Callan, 2003; Ojanen et al., 2005; Rüsseler, Ye, Gerth, Szycik, & Münte, 2018), which is the contrast between the brain activities in response to audiovisual congruent and incongruent stimuli. This is motivated by the fact that the congruency effect is only possible after the unimodal information has been successfully integrated (van Atteveldt, Formisano, Blomert, & Goebel, 2007; van Atteveldt, Formisano, Goebel, & Blomert, 2007). The congruency effect has the advantage of not being sensitive to other common neural activity that is not sensory-specific and therefore holds better statistical criteria. Previous studies (Besle et al., 2004; Cappe, Thut, Romei, & Murray, 2010; Jost et al., 2014; Raji et al., 2000) have shown that the additive effect is associated with the more general audiovisual processing (which includes audiovisual interaction effects in both early and late time window), whereas the congruency comparison is more relevant for the brain interaction of meaningful or already learned audiovisual stimuli (Hocking & Price, 2009).

1.3.2 Brain regions involved in audiovisual processing

Multisensory interaction has been reported at various cortical and subcortical brain regions across species. For example, the superior colliculus (SC) in the midbrain receives auditory, visual, and somatosensory inputs, and its multisensory properties have been documented by numerous studies (Meredith & Stein, 1983; Meredith & Stein, 1986; Perrault et al., 2005; Stein, 1978). Cortical regions, including the superior temporal cortex (Beauchamp, Argall, Bodurka, Duyn, & Martin, 2004; Noesselt et al., 2007; Stevenson & James, 2009), the

intraparietal sulcus (IPS) (Cohen, 2009; Cohen & Andersen, 2004; Molholm et al., 2006) and specific prefrontal regions (Diehl & Romanski, 2014; Macaluso & Driver, 2005) have also been implicated in multisensory integration. In addition, some of the traditionally unisensory regions also showed multisensory properties (Driver & Noesselt, 2008; Ghazanfar & Schroeder, 2006; Kayser, Petkov, Augath, & Logothetis, 2007) or could receive direct inputs from other multisensory (Macaluso, Frith, & Driver, 2000) or unimodal regions (Lakatos, Chen, O'Connell, Mills, & Schroeder, 2007; Murray et al., 2005). Converging evidence suggests the superior temporal cortex (STC) seems to be an important audiovisual integration site in humans. STC has been identified as the primary integration area in studies using different kinds of audiovisual stimuli, for example, audiovisual speech (Calvert, 2001; Calvert, Campbell, & Brammer, 2000; Sekiyama, Kanno, Miura, & Sugita, 2003), audiovisual objects (e.g., tools) (Beauchamp, Argall, et al., 2004; Beauchamp, Lee, et al., 2004) and grapheme-phoneme combinations (Raij et al., 2000; van Atteveldt et al., 2004; van Atteveldt, Roebroek, & Goebel, 2009).

The frontal (including inferior frontal gyrus and premotor) and parietal lobes were also involved during audiovisual processing of speech (Calvert & Campbell, 2003; Ojanen et al., 2005), objects (Hein et al., 2007) and grapheme-phoneme associations (van Atteveldt, Formisano, Goebel, & Blomert, 2007). Recent neuroimaging studies (Doehrmann & Naumer, 2008) suggest potential functional segregation of frontal and temporal cortical networks with the frontal region being more reactive to semantically congruent audiovisual stimuli and the temporal region being more reactive to semantically incongruent audiovisual stimuli. Overall, these findings suggest that the superior temporal cortex (STS/STG) is a general site for integrating learned audiovisual identity information while other regions such as inferior frontal and parietal areas are also involved in specific situations, subserving functions such as top-down control and semantic matching.

1.3.3 Timing of brain activity related to audiovisual processing

fMRI has the advantage of accurately localizing brain regions related to audiovisual integration, but the low temporal resolution of BOLD responses is not suitable for measuring the temporal dynamics of the integration process. Electrophysiological methods such as EEG or MEG could measure brain activity with fine-grained temporal resolution (millisecond) and provide a wide range of temporal and spectral information on the multisensory integration process. In one EEG study (Molholm et al., 2002) using simple auditory tones and visual (a red disk) stimuli, audiovisual integration has been found to occur early (about 40-50 ms) in the right parieto-occipital region, and this integration could affect early visual processing. A similar early audiovisual interaction onset (70-80 ms) was reported in one MEG study (Raij et al., 2010) using simple auditory (noise bursts) and visual stimuli (checkerboards). A suppressive integration effect [$AV < (A + V)$] has been observed in an early time window (50-60 ms after the onset of the stimulus) using simple disc and triangular sound waveform and was

localized in the primary auditory cortex, the primary visual cortex and the posterior superior temporal sulcus (Cappe et al., 2010).

For more complex and language-related audiovisual stimuli such as words, the integration effects seemed to occur later than they do with short and simple audiovisual stimuli. Sensory responses to letters and speech sounds have been reported in one MEG study (Raij et al., 2000) to elicit maximal brain activation in multisensory regions about 200 ms after the onset of audiovisual stimuli. It was followed by the suppressive interaction effects at the time windows of 280–345 ms (in the right temporo-occipito-parietal junction), 380–540 ms (in the left superior temporal sulcus) and 450–535 ms (in the right superior temporal sulcus). Hiragana grapheme–phoneme stimuli were used in one EEG study (Herdman et al., 2006) and a stronger brain oscillation (2–10 Hz) within 250 ms in the left auditory cortex and a weaker brain oscillation (2–16 Hz) at the time window of 250–500 ms in the visual cortices were found in response to the congruent audiovisual stimuli than incongruent audiovisual stimuli. Another ERP study (Jost et al., 2014) discovered that audiovisual suppression effects were 300–324 ms and 480–764 ms for familiar German words, and were 324–384 and 416–756 ms for unfamiliar English words.

1.3.4 Audiovisual integration in alphabetic languages

Existing studies have mostly centered on letter–speech sound processing in alphabetic orthographies such as English (Holloway, van Atteveldt, Blomert, & Ansari, 2015), Finnish (Raij et al., 2000), and Dutch (van Atteveldt et al., 2004; van Atteveldt et al., 2009). Several multisensory brain areas have been identified to show consistent activation patterns during letter–speech sound integration. In particular, the superior temporal cortex has been reported to show heteromodal properties in numerous fMRI studies, with a stronger cortical response to congruent audiovisual stimuli than to incongruent audiovisual (letter–speech sound) stimuli (Blau, van Atteveldt, Formisano, Goebel, & Blomert, 2008; van Atteveldt et al., 2004; van Atteveldt et al., 2009). The bilateral superior temporal cortices were also reported as the major cross-modal integration sites in one MEG study using Finnish letter–speech sound pairs (Raij et al., 2000). Evidence suggests that the multisensory superior temporal cortex could send different feedback projection to the auditory cortex depending on the congruency information processed in STC (van Atteveldt et al., 2004). Furthermore, the audiovisual integration in the superior temporal cortex could arise at a broad range of temporal synchrony between the auditory and visual modalities, whereas the congruency effect in the auditory cortex (planum temporale and Heschl’s sulcus) requires much stricter temporal synchrony (van Atteveldt, Formisano, Blomert, & Goebel, 2007). The congruency effect could also be affected by top-down control mechanisms such as the experiment instructions and task demands (Andersen, Tiippana, & Sams, 2004). For instance, different experimental designs (explicit vs. implicit and active vs. passive) have been shown to modulate the letter–speech sound congruency effect in fMRI (Blau et al., 2008; van Atteveldt, Formisano, Goebel, & Blomert, 2007).

Audiovisual integration has been reported to have an orthographic dependency. For example, the congruency effect was found in the STC in transparent languages such as Finnish (Raij et al., 2000) and Dutch (van Atteveldt et al., 2004). However, in opaque orthography such as English, only a smaller modulation (and in the opposite direction) was found in the brain responses to the less transparent letter–speech sound pairs (Holloway et al., 2015). As discussed in the previous section, the timing of the audiovisual integration effects in alphabetic scripts start relatively late (normally about 200-300 ms after the onset of stimuli) as revealed by time-sensitive EEG/MEG measures, thereby supporting the feedback projection mechanisms (van Atteveldt et al., 2004).

1.3.5 Behavioral correlates of audiovisual integration

Brain activity related to audiovisual processing has been reported to correlate with reading-related cognitive skills. For example, neural activities during a cross-modal rhyme judgment experiment were found to correlate with phonemic awareness for typically developing children but not in children with reading difficulties (RD) (McNorgan, Randazzo-Wagner, & Booth, 2013). Similarly, audiovisual integration in the left STS was correlated with orthographic awareness, word reading ability, and phoneme analysis and elision in typically developing readers (Plewko et al., 2018). Audiovisual integration in temporoparietal reading networks induced by short audiovisual training has been reported to be associated with later reading fluency and therefore shows promising implications for designing early interventions of reading difficulty (Karipidis et al., 2018). Nonetheless, since fMRI has a poor temporal resolution, the above neuroimaging studies were not able to differentiate the underlying sensory and cognitive processes, which might underlie the significant correlations with reading-related cognitive skills.

1.4 Audiovisual learning in the human brain

1.4.1 Letter–speech sound learning

The brain mechanisms of the well-established grapheme–phoneme integration process has been studied in literate adults (Blau et al., 2008; Froyen et al., 2008; Raij et al., 2000; van Atteveldt et al., 2004; van Atteveldt et al., 2009) and in children learning to read (Blau et al., 2010; Froyen et al., 2009; Froyen, Willems, & Blomert, 2011; Žarić et al., 2014). These studies identified brain networks that are consistently activated during letter–speech sound integration days, months, or even years after learning of grapheme–phoneme associations. However, much less is understood about the cognitive processes during the learning of new associations, which is arguably more complex and demanding than the automatic processing of existing associations. The scarcity of cross-modal studies on the learning process in humans is likely due to challenges in studying the

brain mechanisms during multisensory learning since it is very dynamic and involves multiple cognitive components such as sensory processing, multisensory integration, attention, memory formation, and consolidation. In addition, the dynamic nature of learning also brings some methodological challenges in studying learning. For example, brain measurement during relatively stable cognitive processes and certain numbers of stimulus repetition are usually necessary for obtaining a good signal to noise ratio (SNR) using most of the neuroimaging tools.

The grapheme-phoneme learning very likely recruits multiple neurobiological mechanisms and consists of several learning stages: First, auditory and visual sensory inputs are processed in the unisensory auditory and visual cortices where the sensory-specific memory traces are also formed. Then auditory and visual information are integrated and combined into audiovisual objects in multisensory brain regions (e.g., STC) based on the spatial-temporal closeness of the multisensory input (e.g., the coincidence in space and time of the audiovisual stimuli) or other top-down brain mechanisms. For example, during explicit learning, attention is directed to the relevant sensory stimuli and learning cues, which greatly enhance the learning performance through top-down control. The cross-modal audiovisual associations are initially stored in the short-term memory system. The short-term memory of audiovisual associations are consolidated through practice and during sleep (Diekelmann & Born, 2010; Dudai, 2012), and possibly transferred and stored in the neocortex for fast and automatic retrieval (Klinzing, Niethard, & Born, 2019). This is based on the complementary learning systems, which suggest a division of labor with the initial rapid learning in the hippocampus (medial temporal regions) and gradual memory consolidation in the neocortical systems (Davis, Di Betta, Macdonald, & Gaskell, 2009; McClelland, McNaughton, & O'Reilly, 1995). However, fast learning effects that occurred as a rapid form of memory consolidation at the time scale of seconds have also been reported in relation to motor-skill learning (Bönstrup et al., 2019). Such rapid consolidation might also play a role in certain types of sensory learning (Hebscher, Wing, Ryan, & Gilboa, 2019).

1.4.2 Training studies on learning of artificial grapheme-phoneme associations

Artificial letter-speech sound training paradigms that simulate the initial stage of reading acquisition in alphabetic languages could provide interesting insights into the brain mechanisms of letter-speech sound learning. Brain changes related to the learning of cross-modal associations have been reported at various time scales, ranging from minutes (Hämäläinen, Parviainen, Hsu, & Salmelin, 2019; Karipidis et al., 2017) and hours (Brem et al., 2018; Taylor, Rastle, & Davis, 2014) to days (Hashimoto & Sakai, 2004; Karipidis et al., 2018; Madec et al., 2016; Quinn, Taylor, & Davis, 2017; Taylor, Davis, & Rastle, 2017) after the initial training of novel letter-speech sound associations.

Several cortical regions have been identified to be active during the formation of cross-modal associations in previous artificial grapheme-phoneme

training studies. For example, the left posterior inferior temporal gyrus and left parieto-occipital cortex have been reported to show neural plasticity in forming new connections between orthography and phonology when learning novel letters in an early fMRI study (Hashimoto & Sakai, 2004). The parietal brain area is also involved in audiovisual mappings during the early stages of literacy acquisition (Quinn et al., 2017; Taylor et al., 2014). On the other hand, the left vOT seems to receive top-down modulation from the superior temporal gyrus (STG) where phonological recoding processes of newly-learned letters occur, and activation in the left vOT was further correlated with the strength of audiovisual associations in a two-day letter-speech sound training (Madec et al., 2016). Similar changes in the left vOT showed increased N170 responses and vOT activation to newly learned characters after a short artificial character-speech sound training (Brem et al., 2018). These brain changes in the left vOT were also correlated with the training performance and were interpreted as a phonologically driven tuning of N170 and vOT (Pleisch et al., 2019). Furthermore, cross-modal associative learning processes might be affected by modulation of attention to important features seen in the activity of the frontal cortices (Hämäläinen et al., 2019). Interestingly, brain changes related to audiovisual learning was correlated with cognitive skills (Karipidis et al., 2018; Karipidis et al., 2017). For instance, integration effects related to audiovisual learning were found in a distributed brain network after a short grapheme-phoneme training (<30 min) in preschool children (Karipidis et al., 2017) with promising implications in identifying children with reading difficulties and predicting reading outcomes in pre-readers (Karipidis et al., 2018).

Despite the emerging insights from the available literature, to date, there is no comprehensive theoretical model of the cognitive processes and their brain level equivalents that are utilized during grapheme-phoneme learning. It is unclear when and how the audiovisual congruency effect starts to emerge in the multisensory superior temporal cortex and how quickly during training the visual specialization between learned letters starts to differ from unfamiliar letters. In addition, the allocation of attention is essential during explicit learning, yet how attentional processing is modulated by the learning material is still unknown. Finally, brain changes related to the early stages of cross-modal memory consolidation, such as after certain amounts of repetition/practice and after overnight sleep, remain poorly understood.

1.5 Aims of the research

This dissertation aimed to systematically investigate audiovisual processing at different phases of learning to read, therefore corresponding to different levels of automaticity of integration. This includes the very beginning of initial exposure to novel letters, the intermediate level, and the overlearned level. Furthermore, to extend the existing findings on letter-speech sound integration in an

alphabetic language, the audiovisual integration was also examined in a logographic script (Chinese), which also involves the semantic process.

The aim of Study I was to examine the cortical activation to logographic multisensory stimuli using MEG. Earlier studies have reported that the audiovisual integration process is orthographic dependent. For example, a reverse congruency effect ($AVI > AVC$) has been reported in more opaque languages such as English (Holloway et al., 2015) compared to more transparent languages. Still, little is known regarding other non-alphabetic language systems (e.g., Chinese). Chinese characters are syllable-based morphemes that contain meaning. Differences in languages are very likely to be reflected in the audiovisual processing; for example, Chinese character-speech sound integration is expected to involve lexical-semantic processing. Therefore it was hypothesized that the character-speech sound integration would involve both audiovisual congruency effect and semantic processing (N400m-like responses) only in the Chinese group, but not in a Finnish group unaware of the Chinese language. Furthermore, the suppressive interaction [AV vs. (A + V)] was expected to reveal a more general multisensory integration pattern in Chinese and Finnish groups.

In Study II, brain responses to both unisensory (letters or speech sounds) and multisensory (letter-speech sound combinations) stimuli were measured using MEG, with the goal of connecting these brain indices to reading development in children. Earlier studies have found a protracted developmental trajectory of audiovisual integration in children with a cross-modal mismatch negativity experiment (Blomert, 2011; Froyen et al., 2009). Previous studies have reported some interesting correlation patterns between brain responses and cognitive skills: e.g., auditory brain response with phonological and reading skills (Lohvansuu, Hämäläinen, Ervast, Lyytinen, & Leppänen, 2018), the visual response with the reading skills (Brem et al., 2010; Maurer, Blau, Yoncheva, & McCandliss, 2010), and audiovisual integration with reading skills (Blau, van Atteveldt, Ekkebus, Goebel, & Blomert, 2009; Blomert, 2011; Plewko et al., 2018). Study II aimed to explore the level of automaticity of audiovisual integration in children learning to read, and the correlational pattern between the brain responses to auditory (speech sounds), visual (letters), and audiovisual (letter-speech sound combinations) stimuli and children's cognitive skills.

The aim of Study III was to study brain mechanisms during the learning of novel grapheme-phoneme associations and the effect of overnight memory consolidation. During cross-modal associative learning, the auditory and visual inputs had to be integrated and encoded into one audiovisual object, while no such integrative processes were needed in non-associative learning. We expect to see distinctive cognitive processes related to attention and memory encoding in non-associative and associative learning. Furthermore, we hypothesized that the learning of grapheme-phoneme associations would change the corresponding unisensory processing of visually presented novel letters, and elicit congruency effects in multisensory conditions. These effects would be modulated by overnight memory consolidation. The unisensory effects were expected to occur in the occipital and parietal regions, mostly due to the different visual and

attentional processes, and the learning of the phonological representation for the Learnable letters at a relatively late time window around 400 ms based on earlier studies (Dehaene et al., 2010; Quinn et al., 2017; Taylor et al., 2014; Xu, Kolozsvari, Monto, & Hämäläinen, 2018; Xu, Kolozsvári, Oostenveld, Leppänen, & Hämäläinen, 2019). The multisensory congruency effects were expected to be elicited in the posterior superior temporal cortices in the late time window only after the learning of audiovisual associations (van Atteveldt et al., 2004; Wilson, Bautista, & McCarron, 2018; Xu et al., 2019). Finally, learning performance was correlated with cognitive skills linked to reading and working memory to explore the key behavioral factors that interact with multisensory non-associative/associative learning speed.

2 METHODS

2.1 Participants

In Study I, two groups of adult participants were recruited: one group of native Chinese speakers (N = 12) who were studying in Jyväskylä and another group of native Finnish speakers (N = 13). In Study II, participants were 29 Finnish-speaking school children. In Study III, data from 30 native Finnish-speaking adults were used. All participants included in these three studies had no impairment in their hearing ability and normal or corrected-to-normal vision. The participants were also checked for the following exclusion criteria: ADHD, history of head injuries, medications that affect the brain, neurological disorders, delays in language development, or other language-related problems. All three studies were conducted in accordance with the Declaration of Helsinki. Ethical approval was received from the Ethics Committee of the University of Jyväskylä. All participants (also children's parents in Study II) gave their written informed consent prior to the experiment.

2.2 Behavioral measures

In Study II and Study III, a number of cognitive tests were implemented for correlation analysis with brain responses (Study II) and learning performance (Study III). The cognitive tests included the following: subtests from the Wechsler Intelligence Scales for Children Third Edition (Wechsler, 1991) for children above six years and Wechsler Preschool and Primary Scales of Intelligence (Wechsler, 2003) for 6-year-old children in Study II; the Wechsler Adult Intelligence Scales (Wechsler, 2008) for adult participants in Study III. Subtests including digit span (working memory; forward and backward tasks), block design (visuospatial reasoning), and expressive vocabulary were carried out. For the digit span test, a string of numbers is pronounced, and the participants were asked to repeat the

numbers in either forward or backward order. During the block design test, the participants were asked to arrange red and white blocks to form the same design they had been shown earlier by the experimenter. In a more difficult section, the participants were presented with the design in a figure only and were asked to build the same design. During the vocabulary test, the participants were asked to describe the meaning of the word they heard.

Phonological awareness was examined in Study II and Study III with the phonological processing task from NEPSY II (Korkman, Kirk, & Kemp, 2007). During the task, the participant had to first repeat a word and then create a new word using one of the following rules: replace one of the phonemes in the word with another or leave out a phoneme or a syllable. To measure phonological processing and verbal short-term memory skills, the non-word repetition task from the NEPSY I test battery (Korkman, Kirk, & Kemp, 1998) was carried out.

The rapid automatized naming test in Study II and Study III (Denckla & Rudel, 1976) included quickly and accurately naming five pictures of letters or common objects. The letters and objects were in five rows, with each row consisting of 15 objects. The audio sound during this task was recorded and was used to calculate the time (seconds) of the task for the analysis.

Reading tests in Study II and Study III consisted of a standardized word list reading test (Häyrynen, Serenius-Sirve, & Korkman, 1999), in which the score was based on the number of correctly read words in 45 seconds; a non-word list reading task based on Tests of Word Reading Efficiency (Torgesen, Rashotte, & Wagner, 1999), where the score was based on the number of correctly read non-words given time of 45 seconds; a pseudoword text reading task (Eklund, Torppa, Aro, Leppänen, & Lyytinen, 2015), in which the scores were calculated from the number of correctly read pseudowords and the total reading time. A writing to dictation task was carried out in which the participants were asked to write the 20 words they heard on a piece of paper. The score was calculated from the number of correctly written words.

2.3 Stimuli and task

In Study I, six Simplified Chinese characters and their associated flat tone speech sounds (1. 酷: ku; 2. 普: pu; 3. 兔: tu; 4. 步: bu; 5. 都: du; 6. 谷: gu) were used as the audiovisual stimuli. Four types of stimuli, including unisensory auditory (A), unisensory visual (V), audiovisual incongruent (AVI), and congruent (AVC) were presented randomly during the experiment. In order to keep the participants' attention equally on the inputs from auditory and visual modalities, they were asked to do a two-modality one-back working memory task.

In Study II, eight Finnish letters (A, Ä, E, I, O, Ö, U and Y) and their associated phonemes ([a], [æ], [e], [i], [o], [ø], [u] and [y]) were used as audiovisual stimuli. A child-friendly experimental design was used for Study II in which the theme was about the forest adventure story of a Finnish cartoon character. The children pressed a button if an animal picture was shown on the

screen or an animal sound was played among the random presentations of A, V, AVI, and AVC trials. Similar detection tasks that require the participants to relate to the audiovisual information explicitly were used in earlier studies (Blau et al., 2010; Raji et al., 2000) on audiovisual integration in adults and children.

In Study III, the visual stimuli consisted of 12 Georgian letters (ჲ, ჳ, ჴ, ჵ, ჶ, ჷ, ჸ, ჹ, ჺ, ჻, ჼ, ჽ, ჾ, ჿ). Auditory stimuli consisted of 12 Finnish phonemes ([a], [ä], [e], [t], [s], [k], [o], [ö], [i], [p], [v], [d]). The auditory and visual stimuli were divided into two sets with six audiovisual pairs in each set. One of the two audiovisual stimulus sets was used as the Learnable set in which different learning cues (✓ for congruent pairs [AVC] and X for incongruent pairs [AVI]) were presented after the simultaneous presentation of audiovisual stimuli. The other audiovisual stimuli set was used as the Control set, in which the feedback was always ✘ after the audiovisual stimuli (AVX). The audiovisual learning experiment consisted of 12 alternating training and testing blocks on the first day and six training and testing blocks on the second day.

2.4 MEG and MRI data acquisition

Magnetoencephalography data were collected with the Elekta Neuromag® TRIUX™ system (Elekta AB, Stockholm, Sweden) in a room with magnetic shielding and sound attenuation at the University of Jyväskylä. A sampling rate of 1000 Hz and an online band-pass filter at the frequency of 0.1-330 Hz were used in data acquisition settings. The head position with reference to the sensor arrays within the MEG helmet was continuously traced using five digitized head position indicator (HPI) coils, of which three were taped on the forehead and one behind each ear. The head coordinate system was defined by three anatomic landmarks, including left and right preauricular points and the nasion. The anatomical landmarks, the position of the HPI coils, and the head shape (>100 points evenly distributed over the scalp) were digitally recorded using the Polhemus tracking systems (Polhemus, Colchester, VT, United States) before the MEG experiment. In order to record the electrooculogram (EOG), two electrodes were attached diagonally with one slightly below the left eye and one slightly above the right eye, and one additional ground electrode was attached to the collarbone. The MEG data were acquired in an upright gantry position (68°), with participants sitting comfortably on a chair.

In Study II, structural magnetic resonance images (MRI) were acquired from Synlab Jyväskylä, a private company specialized in MRI services. T1-weighted 3D-SE images were acquired on a GE 1.5 T MRI scanner (GoldSeal Signa HDxt) with a standard head coil and using the following parameters: TR/TE = 540/10 ms, sagittal orientation, matrix size = 256 × 256, flip angle = 90°, slice thickness = 1.2 mm.

2.5 Data analysis

Common MEG data analysis steps between the studies included: first, pre-processing with Maxfilter (Version 3.0) to remove the external noise interference and to compensate for head movement during the recording using a movement-compensated temporal signal-space separation (tSSS) method (Taulu & Simola, 2006). MEG channels were checked manually to remove bad ones from the Maxfilter. These bad channels were then reconstructed after the Maxfilter.

Second, data were then analyzed with the open source toolbox MNE Python (Gramfort et al., 2013) and FieldTrip (Oostenveld, Fries, Maris, & Schoffelen, 2011). A 40 Hz low-pass filter (zero-phase FIR filter design using the window method) was applied to the MEG data. Fast ICA (Hyvärinen, 1999) was then applied to remove eye movement-related or cardiac artifacts. After applying ICA, data were segmented into epochs with 200 ms (Study I and II) or 150 ms (Study III) prior to and 1000 ms after the stimulus onset. Then epochs were checked manually, and bad epochs were removed from further analysis. Baseline correction was implemented by removing the mean response before the stimulus onset from the whole epoch.

Individual MRIs from Study II were analyzed using Freesurfer (RRID: SCR_001847, v5.3.0, Martinos Center for Biomedical Imaging, Charlestown, MA, United States) to construct the cortical surface for source localization. Individual MRIs were not available for the adult participants in Study I and Study III, and therefore the FSAverage brain template from Freesurfer (RRID: SCR_001847, v5.3.0, Martinos Center for Biomedical Imaging, Charlestown, MA, United States) was used. Coregistration was done between the digitized head surface and the brain template with 3-parameter scaling.

Cortically constrained minimum-norm source estimate (MNE) with depth-weighting ($p = 0.8$) (Hämäläinen & Ilmoniemi, 1994; Lin et al., 2006) was used for source analysis. A one-layer boundary element model (BEM) derived from the inner skull surface was applied for the forward modeling. The pre-stimulus baseline data pooled from all conditions were used for the estimation of the noise covariance matrix. For each of the current dipoles in the source space, the source amplitude values were calculated using the vector norm. In Study II, the Desikan-Killiany Atlas was used to calculate the mean source amplitude values within each of the 68 defined brain regions (Desikan et al., 2006). In Study I and Study III, the dynamic statistical parametric maps (dSPM) (Dale et al., 2000) were applied for noise normalization after the MNE estimation.

In all three studies, interaction ([A + V vs. AV]) and congruency ([AVC vs. AVI]) effects were used for investigating the audiovisual processing. In Study II, the source-level brain activation of visual (P1m and N170m) and auditory (N1m, N2m, and late sustained component) event-related fields (ERF) components were extracted for the regression analysis with children's cognitive skills. In Study III, a learning index for each audiovisual stimulus was calculated based on the performance in the testing blocks. Based on the learning progress, the

participants acquired the letter–speech sound association adequately after about four blocks of successful learning. The MEG data for Day 1 were therefore split over three learning stages (learning index = 0, 1–4, and >4) for the audiovisual conditions in learning and testing conditions separately. For Day 2, the MEG data were averaged together, since the participants had already learned all the audiovisual pairs. For the different learning cues, we postulated that the participants were paying attention to them before learning and immediately following the first few successful learning trials. Therefore, MEG data were split into the following three parts for comparing the learning cues: learning index 0–4, learning index >4 on Day 1 and all the data on Day 2. The unisensory auditory and visual responses (for Learnable vs. Control comparison), as well as the brain responses to three different learning cues, were calculated separately for different learning stages in two days in Study III.

2.6 Statistical analysis

In Study I, cluster-based (spatiotemporal) nonparametric tests (Maris & Oostenveld, 2007) were conducted for testing the interaction ([A + V vs. AV]) and congruency ([AVC vs. AVI]) effects within Chinese and Finnish groups separately at both sensor and source levels. Combined gradiometer data were used in the sensor-level statistical analysis, which was implemented in the Fieldtrip toolbox. Similar statistical tests were carried out at the source level using the MNE Python toolbox.

In Study II, partial correlation (control for the effect of age) in SPSS (version 24, IBM Corp., Armonk, NY, United States) was used to examine the relationship between the children’s cognitive skills and the brain activities (mean source amplitudes and peak latencies of brain sensory responses from all four conditions). Based on the results from the significant partial correlations, a linear regression model was constructed in SPSS with brain activities as independent variables and the children’s cognitive skills as dependent variables. The age of the participants was entered into the regression model followed by the brain responses (stepwise method: age->auditory/visual-> audiovisual) to explore the unique variance explained by each independent variable. Temporal cluster-based nonparametric permutation tests implemented in the Mass Univariate ERP Toolbox (Groppe, Urbach, & Kutas, 2011) were used for testing the audiovisual interaction ([A + V vs. AV]) and congruency ([AVC vs. AVI]) effects at the source level (68 brain regions defined by the Desikan-Killiany Atlas). For brain regions that demonstrated significant ($p < 0.05$) interaction or congruency effects, partial correlations (controlling for the effect of age) were computed between cognitive scores and multisensory brain activations in these brain areas by taking the mean values from the time window of the clusters exceeding the randomization distribution under H_0 . A data-driven approach (whole brain with broad time window: 0–1000 ms) was used due to the small number of studies examining

these effects in children compared to the clearly defined hypothesis for the obligatory sensory responses.

In Study III, region of interest (ROI) analysis was used for comparing AV congruency effect in a 3 (congruency: AVC, AVI, AVX) \times 2 (hemisphere: left, right) analysis of variance (repeated measures ANOVA in SPSS) model. Based on earlier literature (Karipidis et al., 2017; Raji et al., 2000; Xu et al., 2019) brain dSPM source waveforms of multisensory responses (500ms to 800ms after stimulus onset) were extracted from the left and right bank of the posterior superior temporal sulcus (pSTS, label: “bankssts”) (Beauchamp, Argall, et al., 2004; Blomert, 2011; Calvert et al., 2001; van Atteveldt et al., 2009; Xu et al., 2019) as defined by the Desikan-Killiany Atlas (Desikan et al., 2006). Cluster-based (spatiotemporal) permutation tests (Maris & Oostenveld, 2007) were used for comparing Learnable and Control auditory, visual, and audiovisual interaction brain activations from the linear regression analysis based on the additive model using MNE Python. Brain responses to different learning cues (“YES”: ✓; “NO”: X; “UNKNOWN”: ☒) were also compared in pairs using the spatiotemporal cluster-based permutation tests. We did not have a clear hypothesis on the time and location of this effect because of insufficient evidence from earlier studies; therefore, a wider time window and whole-brain approach were used for the spatiotemporal cluster-based permutation tests. Finally, to explore how much variance of the reading-related cognitive scores could be explained by the learning speed of Learnable and Control stimuli, correlation analysis (Pearson’s correlation coefficients) was carried out between the individual learning speed (average learning index of all Learnable and Control stimuli pairs in the twelfth block) on Day 1 and all the cognitive test scores. The false discovery rate (FDR) was applied to correct the p-values in the correlation analysis for the number of tests (Benjamini & Hochberg, 1995).

TABLE 1 Summary of methods in all three studies.

Study	Participants	Age (mean \pm SD)	Measure	Experiment	Statistics
I	Chinese: N = 12 Finnish: N = 13	Chinese: 24.36 \pm 366 Finnish: 24.31 \pm 2.06	MEG	Audiovisual integration (Chinese)	Spatiotemporal cluster-based permutation tests
II	Finish children: N = 29	8.17 \pm 1.05	MEG, MRI, Cognitive tests	Audiovisual integration (Finnish)	Regression analysis, Temporal cluster-based permutation tests
III	Finish adults: N = 30	24.33 \pm 3.50	MEG, Cognitive tests	Audiovisual learning (Georgian)	ANOVA, Spatiotemporal cluster-based permutation tests

3 RESULTS

3.1 Study I

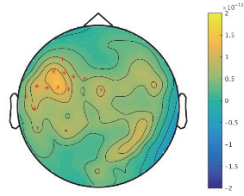
In Study I, the spatiotemporal dynamics of brain activation in response to logographic multisensory (auditory and/or visual) stimuli were examined by applying interaction and congruency contrasts in Chinese and Finnish groups.

Suppression effects [$AV < (A + V)$] were observed in both samples (Chinese and Finnish groups) at the sensor and the source levels but with a left-lateralized effect (left temporal and frontal) in the Chinese group and a right-lateralized (right parietal-occipital) effect in the Finnish group. As expected, the congruency effect was only found in the Chinese group at both the sensor and the source level (left frontal and temporal) since only the Chinese participants had knowledge of the correct audiovisual associations. Overall, the sensor- and source-level statistical results showed converging patterns regarding the time window and spatial regions of clusters exceeding the threshold of randomization distribution under H_0 . Details of the significant effects are reported in Table 2 and Figure 1.

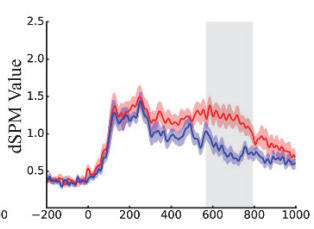
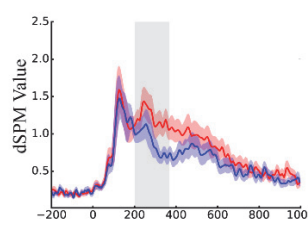
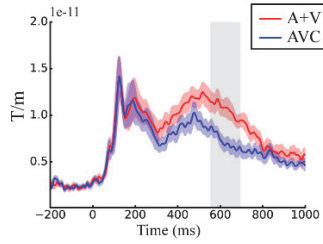
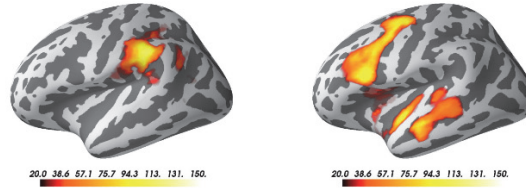
Suppression Effect

Chinese

Sensor

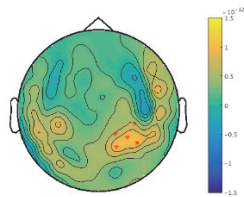


Source

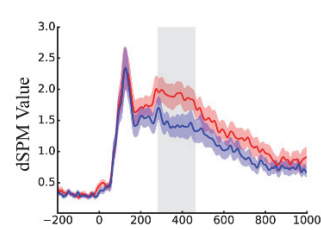
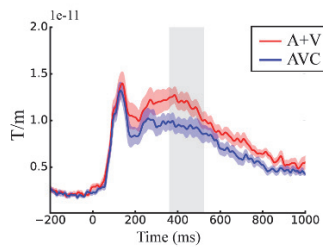
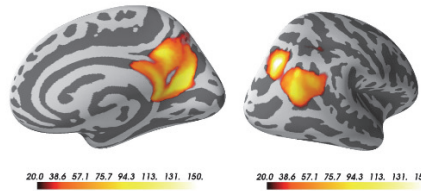


Finnish

Sensor



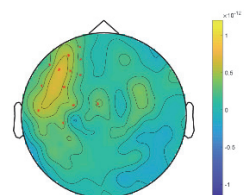
Source



Congruency Effect

Chinese

Sensor



Source

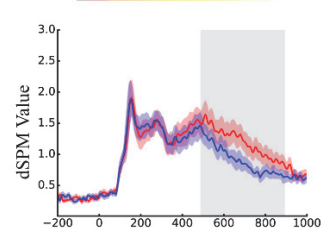
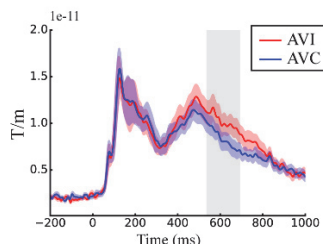
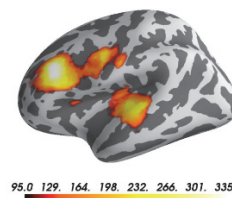


FIGURE 1 Statistical results of suppression and congruency effects at the sensor and source levels for the Chinese and Finnish groups. For the sensor-level statistical results, the clusters exceeding the randomization distribution under H0 are highlighted by red dots representing those channels in the sensor space. The clusters are overlaid on the sensory topography of the difference contrast extracted from the time window of clusters. For the source level, the clusters exceeding the randomization distribution under H0 are highlighted by the yellow and red coloring on the cortical surfaces. The brightness of the cluster is scaled by the temporal duration of the cluster in the source space. In addition, average evoked responses from the channels of the cluster are plotted beneath the sensor space results, and the source waveform (dSPM value) extracted from the clusters is plotted beneath for the source space results. The red and blue shaded area defines the standard error of the mean, and the gray shaded area indicates the time window of the cluster.

TABLE 2 Summary of the clusters exceeding the randomization distribution under H0 for suppression and congruency effects at sensor and source levels in the Chinese (N = 12) and the Finnish (N = 13) groups.

Effect	Level	Group	Cluster number	Time window (ms)	Region	p-value
Suppression effect	Sensor	Chinese	1	557-692	Left temporal & frontal	0.002
		Finnish	1	363-520	Right parietal-occipital	0.006
	Source	Chinese	1	205-365	Left angular & supramarginal gyri	0.01
			2	575-800	Left temporal & frontal	0.001
		Finnish	1	285-460	Right parietal-occipital	0.003
Congruency effect	Sensor	Chinese	1	538-690	Left frontal & temporal	0.01
	Source	Chinese	1	490-890	Left frontal & temporal	0.008

3.2 Study II

In Study II, both unisensory (A and V) and multisensory (AVC and AVI) brain responses to Finnish letters and corresponding phonemes were measured using MEG and were correlated with children's reading-related cognitive skills after controlling for the effect of age. The age effect was controlled to investigate the reading development independent of brain maturation. Multisensory interaction and congruency effect were also examined, and significant brain indices of audiovisual integration were further correlated with cognitive abilities.

Partial correlation analysis revealed a rather consistent correlation pattern between auditory and audiovisual responses (N1m, N2m, and late component) in auditory cortices and phonological processing and rapid naming speed (letters). The audiovisual responses (both AVC and AVI) in the left fusiform areas (N170m) were significantly associated with phonological processing skills.

All significant correlation results are shown in Table 3. The correlation pattern seemed to suggest a substantial overlap in children's brain responses in terms of relationships with cognitive skills. To disentangle these overlapping variances between these brain responses, more specifically to see if there is independent contribution from the multisensory responses, a linear regression model was used with phonological processing and rapid naming as dependent variables. The stepwise method was applied: age was put to the model first followed by the unisensory brain responses (auditory/visual) and then the multisensory brain responses (AVC and AVI). The regression results suggested that auditory late component was the only significant predictor of phonological processing and rapid naming skill of letters. Therefore, it can be concluded that a larger auditory response (late sustained component) to the phoneme stimuli is associated with faster rapid naming skills and better phonological processing abilities in children.

No significant congruency effects were found by the cluster-based permutation test ($p > 0.05$). An audiovisual interaction effect ($[AVC < (A + V)]$) was found in distributed parietal and temporal brain regions (see Figure 2). The clusters exceeding the randomization distribution under H_0 and their time windows were as follows: left (317–499 ms) inferior parietal, left (391–585 ms) supramarginal, right (315–818 ms) inferior parietal, right (306–797 ms) supramarginal, right (271–529 ms) precuneus, right (551–755 ms) postcentral, right superior (535–827 ms) temporal and right middle (346–749 ms) temporal cortices. Furthermore, the source amplitudes of significant audiovisual interaction effects $[(A + V) - AVC]$ in the above eight regions were checked for correlations (partial correlation; controlling for the effect of age) with cognitive skills. Audiovisual interaction in the right precuneus and inferior parietal region was significantly correlated with phonological processing skills, whereas audiovisual interaction in the right precuneus and left supramarginal region was correlated with rapid naming speed of letters. Furthermore, audiovisual interaction in the right precuneus and right temporal (superior and middle temporal) regions was correlated with reading (word list, non-word list, and non-word text reading) and writing skills, respectively.

TABLE 3 Significant partial correlations between sensor brain responses and reading-related cognitive skills (controlling for the effect of age).

Cognitive skills	Auditory Cortex					Visual cortex				
	A		AVC		AVI		AVC		AVI	
	Com (Hemi)	r	Com (Hemi)	r	Com (Hemi)	r	Com (Hemi)	r	Com (Hemi)	r
Phonological processing	N1m (R)	0.384*	N2m (L)	0.420*	N1m (L)	0.422*	N170 (L)	0.404*	N170 (L)	0.427*
	N2m (L)	0.454*	LC (L)	0.506**	LC (L)	0.441*				
	LC (L)	0.499**	LC (R)	0.448*	LC (R)	0.456*				
	LC (R)	0.472*								
RAN Letter	LC (L)	-0.399*	LC (L)	-0.412*	LC (L)	-0.381*				
RAN Object			LC (L)	-0.405*	LC (L)	-0.394*				
Non-word list reading			N1m (R)	0.390*						
			LC (L)	0.395*						

Note. A = auditory, AVC = audiovisual congruent, AVI = audiovisual incongruent, Com = component, Hemi = hemisphere, LC = late component, r = correlation coefficient, L = left hemisphere, R = right hemisphere, *p < 0.05, **p < 0.01.

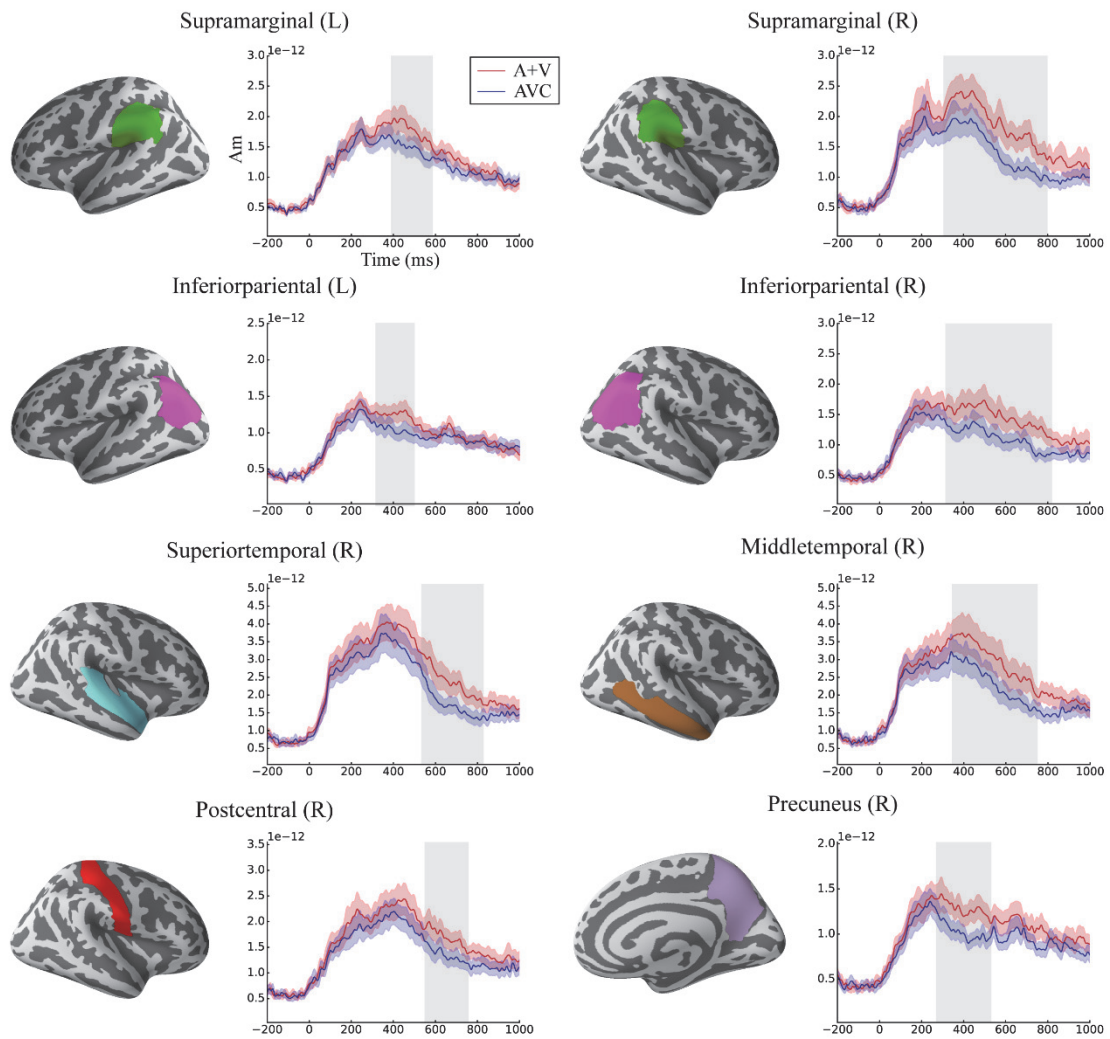


FIGURE 2 Brain regions that showed significant suppressive interaction effects ($A + V > AVC$). The eight temporoparietal brain regions (defined in the Desikan-Killiany atlas) showing significant effects are highlighted on the cortical surfaces (in the left panels). The right panels show the average source waveform (MNE estimation) extracted from the brain regions with significant differences. The red and blue shadings define the standard error of the mean, and the gray shadings represent the time window of the cluster.

3.3 Study III

In Study III, we investigated the neural mechanisms during the learning of novel grapheme-phoneme associations and the effect of overnight memory consolidation. The learning progress was tracked on a trial-by-trial basis during training on two consecutive days and was used to identify and segment different learning stages.

3.3.1 Congruency effects in the pSTS

Repeated-measures ANOVA revealed significant main effects of the congruency on Day 1 only after learning of letter–speech sound associations in the training blocks (learning index >4: $F(2, 52) = 4.81, p = 0.017$) and in the testing blocks (learning index 1–4: $F(2, 58) = 4.37, p = 0.022$; learning index >4: $F(2, 54) = 4.43, p = 0.022$), as well as on Day 2 ($F(2, 58) = 3.82, p = 0.034$) during the training blocks. Post-hoc t-tests indicated that dSPM activation to the Control audiovisual stimuli (AVX) was significantly lower ($p < 0.05$) than to the Learnable audiovisual stimuli (AVC and AVI) in the Day 1 training blocks (when learning index: >4) and to the audiovisual congruent stimuli (AVC) in the Day 2 training blocks. During the testing blocks when the learning index was 1–4, the incongruent audiovisual stimuli (AVI) elicited significantly higher ($p < 0.05$) activation than the Control audiovisual stimuli (AVX). The Learnable congruent audiovisual stimuli (AVC) elicited significantly higher ($p < 0.05$) activation than the Learnable incongruent audiovisual stimuli (AVI) and the Control stimuli (AVX) in the Day 1 testing blocks when the learning index was greater than 4. In addition, there was a hemisphere main effect ($F(1, 29) = 7.48, p = 0.011$) with higher dSPM activation in the right hemisphere than the left hemisphere during the training blocks on Day 1 at the stage when the learning index was 1–4.

3.3.2 Cortical responses to unimodal stimuli and audiovisual interaction (Learnable vs. Control)

No significant differences were found for all the comparisons on Day 1 and Day 2 for auditory responses (Learnable vs. Control). For the visual response, significant differences were found between learnable and control conditions when the learning index is greater than 4 ($p = 0.002, 455\text{--}795$ ms, left parietal and occipital regions) on Day 1 and Day 2 ($p = 0.001, 380\text{--}795$ ms, left parietal and occipital regions). The cross-modal learning effects were tested for the audiovisual interaction by comparing the audiovisual interaction brain activations of the Learnable and Control stimuli from the time window of 500 ms to 800 ms after the stimulus onset using spatiotemporal cluster-based permutation statistics. The statistical tests were carried out for the three different learning stages (Learning index = 0, 1–4, >4 respectively) on Day 1 and the learned stage on Day 2. There was a significant difference ($p = 0.019, 500\text{--}680$ ms, left parietal region) when at the stage when the learning index was 1–4 on Day 1. As expected, the above results suggest that no brain activity changes were observed before learning of the cross-modal association (learning index = 0). Changes only started to occur after the successful learning of audiovisual associations (learning index > 0).

3.3.3 Cortical responses to different learning cues

The brain activities following the three different learning cues were compared in pairs using the spatiotemporal cluster-based permutation tests in the time

window of 100 ms to 800 ms for Day 1 (Learning index = 0–4, >4) and Day 2. There were significant differences between the three different learning cues when the learning index was between 0–4 on Day 1. Two clusters exceeding the randomization distribution under H0 were found for the ✓ vs. X comparison, one ($p = 0.012$) in the left temporal regions in the time window of 300–490 ms and another ($p = 0.016$) in the right temporal regions in the time window of 295–550 ms. Two clusters exceeding the randomization distribution under H0 were found for the ▣ vs. X comparison, one ($p = 0.008$) in the left temporal regions in the time window of 360–730 ms and another ($p = 0.036$) in the right temporal regions in the time window of 355–785 ms. Two clusters exceeding the randomization distribution under H0 were found for the ✓ vs. ▣ comparison, one ($p = 0.040$) in the left temporal regions in the time window of 400–780 ms and another ($p = 0.037$) in the right temporal regions in the time window of 245–455 ms. In addition, there was a significant difference for the ▣ vs. X comparison ($p = 0.029$, 300–740 ms, left temporal and occipital regions) when the learning index was greater than 4 on Day 1. No significant differences were found between three different learning cues on Day 2. The results regarding the Learnable vs. Control comparisons (auditory, visual, audiovisual interaction) and contrasts between the brain responses to different learning cues are summarized in Figure 3.

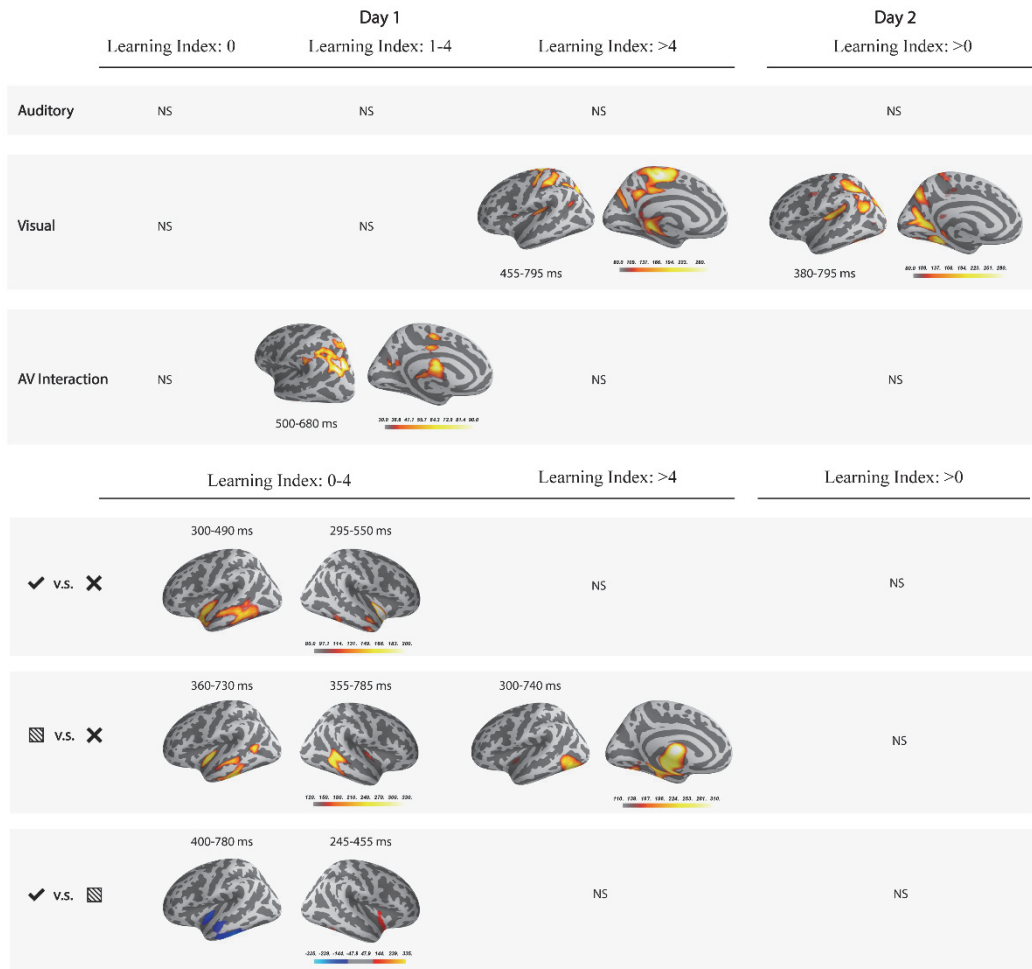


FIGURE 3 The spatiotemporal cluster-based statistical results for the Learnable vs. Control comparisons (auditory, visual, audiovisual interaction) and contrasts of the brain responses to different learning cues at different learning stages in two days. MEG data were split into the following three parts for the Learnable vs. Control comparisons (auditory, visual, audiovisual interaction): learning index 0, learning index 1-4, learning index >4 on Day 1, and all the data on Day 2. MEG data were split into the following three parts for contrasts of the brain responses to different learning cues: learning index 0-4, learning index > 4 on Day 1, and all the data on Day 2. The brain activities were compared using the spatiotemporal cluster-based permutation tests. The cluster exceeding the randomization distribution under H0 was represented by red coloring on the cortical surface, and the temporal duration of the whole cluster was marked in the figure above the cortical surface. The brightness of the color on the cortical surface was scaled by the temporal duration of the cluster. A warm color means the difference is significantly greater than zero, and a cold color means the difference is significantly smaller than zero. Non-significant results are marked with NS.

3.3.4 Correlations between cognitive skills and learning speed

Correlation analysis was carried out between learning speed (of the Learnable and Control stimuli) and cognitive test scores. After FDR correction, only the learning speed of the Control stimuli was significantly correlated with the time spent on RAN objects (FDR-corrected $p = 0.0002$).

4 DISCUSSION

This dissertation aimed to investigate the brain activity changes related to the learning of audiovisual associations in reading. These include the dynamic changes in cortical activity during the learning (Study III), at early stages (Study II) after the learning of letter-speech sound associations and audiovisual processing of meaningful character-speech sound combinations (Study I). Study I investigated the audiovisual integration process in a logographic language, Chinese, in which a character is associated with a syllable sound and a meaning. One group of native Chinese speakers and another group of native Finnish speakers (control group) participated in the audiovisual MEG experiment with Chinese characters and speech sounds as stimuli. For the Finnish group, only a suppression effect was found in the right parietal and occipital cortices, which is probably related to the general audiovisual process of unfamiliar audiovisual stimuli. Both audiovisual suppression and congruency effects were found in the Chinese group in the left superior temporal and left inferior frontal regions, which reflected the effect of learning character-speech sound associations. Study II examined the brain activities of unisensory (auditory/visual) and audiovisual processing in a group of Finnish children learning to read with the aim of linking these cortical unisensory and audiovisual responses to their reading-related cognitive skills. Regression analysis showed that from the brain measures, the auditory late response around 400 ms showed the largest association with phonological processing and rapid automatized naming abilities. In addition, the audiovisual integration effect was most pronounced in the left and right temporoparietal regions, and activations in several of these regions were correlated with children's reading and writing skills. Study III aimed to investigate the cortical mechanisms for supporting letter-speech sound learning, particularly the brain dynamics during the learning of grapheme-phoneme associations. Dynamic changes were observed in brain responses related to multisensory processing and the visual letter processing when grapheme-phoneme associations were learned and after overnight memory consolidation. Overall, the cross-modal learning process changes brain activity in a large network of brain regions, including the superior temporal cortex and the dorsal

(parietal) pathway. Most interestingly, after the presentation of the cross-modal relationship (from learning cues), the middle- and inferior- temporal regions seemed to be engaged in multisensory memory encoding processes.

4.1 Audiovisual integration in logographic languages

In Study I, we investigated audiovisual integration in logographic scripts (Chinese) in one group of native Chinese speakers and another group of native Finnish speakers. The suppressive effect [$AV < (A + V)$] showed different patterns in the two groups: it was left-lateralized for the Chinese group, which possibly reflecting automatic audiovisual processes of learned character-speech sounds, and right-lateralized for the Finnish group when processing novel Chinese audiovisual stimuli. In addition, the congruency effect was only found in the Chinese group in the left superior temporal cortex and Broca's area in a late time window (500–800 ms).

The suppressive effect was found mainly in the left angular/supramarginal, inferior frontal, and temporal brain areas for the Chinese group, which indicated that the left-hemispheric language network was activated during the processing of well-learned audiovisual (Chinese character and speech sound) associations. The suppression effect in the left angular and supramarginal gyri was found relatively early at the time window of about 200–350 ms after the stimulus onset. The left angular and supramarginal gyri have been identified as heteromodal areas which link orthographic representations of symbols from the occipital region to phonological coding in the superior temporal cortex (Price, 2000; Pugh, Mencl, Shaywitz, et al., 2000; Schlaggar & McCandliss, 2007) and also possibly to the inferior frontal gyrus through a feedforward pathway (Simos, Rezaie, Fletcher, & Papanicolaou, 2013). The suppressive effect in the left superior and middle temporal regions in the late time window (around 550–800 ms) was consistent with results from earlier MEG/EEG (Calvert et al., 2001; Raij et al., 2000) and fMRI (van Atteveldt et al., 2004; van Atteveldt et al., 2009) studies using alphabetic letters. The suppressive effect in Broca's area could be related to the extra semantic processing of Chinese characters (Kuo et al., 2001; Tan et al., 2001; Wang et al., 2008), which was not needed for letter-speech sound processing in alphabetic languages. The audiovisual suppression effect could be seen as the optimization of brain networks as a result of language learning (Raij et al., 2000) for the Chinese participants.

The suppressive effect was found in a comparatively earlier time window (285–460 ms) in the right inferior parietal and occipital area in the Finnish group. For the Finnish participants who had never learned Chinese, they most probably process the unfamiliar audiovisual information by paying more attention to the visual features (Calvert et al., 2001; Madec et al., 2016), for example by analyzing the spatial configuration of varying strokes within the character to be able to process the audiovisual stimuli. The parietal region has been known to have multisensory properties (Ahveninen et al., 2006; Bremmer et al., 2001; Cohen,

2009; Grunewald, Linden, & Andersen, 1999; Lewis, Beauchamp, & DeYoe, 2000). Converging evidence also suggests that the right parietal region is related to the perception of multisensory stimuli that do not have any long-term memory representations from unisensory modalities (Kamke, Vieth, Cottrell, & Mattingley, 2012). Therefore, the suppressive effect found in the right parietal region in the Finnish group most likely reflects a basic audiovisual processing of novel symbols (Calvert, 2001; Cappe et al., 2010; Molholm et al., 2002) during a working memory task or attention-related process in the distributed frontoparietal cortical networks (Scolari, Seidl-Rathkopf, & Kastner, 2015; Shulman et al., 2010; Yantis, 2008).

The congruency (AVC vs. AVI) effect was only significant in the Chinese group, which were mainly distributed in the left superior temporal cortex (including Heschl's gyrus), left inferior frontal cortex and also parts of the insula. The congruency effect found in the superior temporal cortex matches previous studies (Blau et al., 2008; Jost et al., 2014; Raji et al., 2000; van Atteveldt et al., 2004) on letter-speech sound integration in alphabetic languages. According to the functional neuroanatomical model proposed by (van Atteveldt et al., 2009), the visual and auditory inputs are integrated in the multisensory STS/STG regions and then possibly projected to the auditory cortex through a feedback pathway. The activation of this feedback projection mechanism was supported by the relatively late time window (around 500–800 ms after the stimulus onset) of the congruency effect in Study I. Earlier studies using transparent languages such as Finnish (Raji et al., 2000) and Dutch (van Atteveldt et al., 2004) found that the brain responses in the auditory association cortex were enhanced by congruent letters-speech sound pairs and suppressed by incongruent letters-speech sound pairs. However, the direction of congruency was reversed for opaque languages such as English, in which incongruent letter-speech sound pairs elicited even stronger activation in the superior temporal cortex than congruent ones (Holloway et al., 2015). The results from Study I further suggested that the congruency is adaptive to different types of scripts showing a reversed direction (AVI>AVC) in the superior temporal areas in logographic scripts compared to earlier studies on alphabetic language (Raji et al., 2000; van Atteveldt et al., 2004). A comparable congruency effect was observed for German words, which was related to lexical-semantic processing of words (Jost et al., 2014).

The left superior temporal cortex and left frontal cortex showed similar effects in both suppressive and congruency comparisons in Study I. This is consistent with previous studies on Chinese characters, which showed that left temporal cortices and frontal regions (BA 9 and BA 47) are important for the character reading process (Tan et al., 2001; Tan et al., 2000). For example, the left middle frontal cortex (BA 9) was indicated to be involved in the visuospatial analysis and semantic processing of Chinese characters (Tan et al., 2001). The inferior frontal cortex has been reported repeatedly to be activated specifically by semantically incongruent audiovisual stimuli, which was associated with increased demands of cognitive control, including working memory processes and semantic retrieval (Doehrmann & Naumer, 2008; Martin & Chao, 2001;

Wagner, Paré-Blagoev, Clark, & Poldrack, 2001). Taken together, the left superior temporal cortex seems to be a common hub in the brain network of audiovisual processing across languages. Furthermore, the left inferior frontal cortex is recruited for the additional semantic process during audiovisual integration in logographic scripts (van Atteveldt & Ansari, 2014; Wu, Ho, & Chen, 2012;).

4.2 Audiovisual processing and its behavioral correlates in children

In Study II, brain-evoked responses to auditory, visual, and audiovisual stimuli of Finnish letters and speech sounds were extracted and correlated with cognitive skills in children learning to read (6-11 years). The results revealed an interesting correlation pattern: auditory responses, especially the late sustained response, were significantly correlated with phonological skills; the visual N170 response from the left fusiform gyrus was also correlated with phonological skills in the audiovisual condition. Furthermore, audiovisual integration effects indexed by the suppressive interaction [$AV < (A + V)$] were found in the temporoparietal regions and contributed independently to reading skills. The congruency effect was not significant, therefore indicating less automatized LSS integration in children learning to read.

The auditory responses to speech sounds had major peaks around 100 ms (P1/N1) and 250 ms (N2), which were typical in children, as reported by earlier studies on the maturation of auditory evoked responses (Parviainen, Helenius, Poskiparta, Niemi, & Salmelin, 2011; Wunderlich, Cone-Wesson, & Shepherd, 2006). The N2 response was followed by another late negative peak around 400 ms (Čeponienė et al., 2001; Čeponienė, Torki, Alku, Koyama, & Townsend, 2008; Szymanski, Rowley, & Roberts, 1999) and is related to phonological processing (Bann & Herdman, 2016; Kuuluvainen, Leminen, & Kujala, 2016; Stevens, McIlraith, Rusk, Niermeyer, & Waller, 2013). In Study II, the auditory N1m, N2m, and late component were found to be correlated with phonological processing skills. This was consistent with earlier studies (Bonte & Blomert, 2004; Hämäläinen, Lohvansuu, Ervast, & Leppänen, 2015; Lohvansuu et al., 2014; Parviainen et al., 2011), which linked auditory N1, N2 and late component to reading and reading-related skills in children. Further regression analysis revealed that among the auditory components in the auditory and audiovisual conditions, the left auditory late response was the driving force and explained unique variance in the correlation with phonological skills. In addition, the left auditory late response amplitude was also significantly correlated with rapid naming ability; similar results have been reported in earlier research (Kuuluvainen et al., 2016).

For the visual modality, the only significant brain-behavior correlation was found between the N170 responses in the left fusiform area under both audiovisual (AVC and AVI) conditions and phonological skills. This was

consistent with findings in another MEG study (Parviainen, Helenius, Poskiparta, Niemi, & Salmelin, 2006) in which letter-string-sensitive activation in the occipitotemporal area was also correlated with phonological abilities in children. The N170 response is an electrophysiological index of brain specialization for letter string or word processing (Maurer, Brem, Bucher, & Brandeis, 2005) and was functionally localized in the so-called visual word form area (VWFA) in the left occipitotemporal area (Cohen et al., 2000; Dehaene & Cohen, 2011; Dehaene, Le Clec'H, Poline, Le Bihan, & Cohen, 2002). Evidence has suggested that the emergence of the left-lateralized N170 response is at least partly driven by an automatic connection between orthographic and phonological systems (Maurer et al., 2010; McCandliss & Noble, 2003). Therefore, the significant correlation between left fusiform response and phonological skills under audiovisual conditions suggested a top-down feedback modulation of the left ventral occipital area from the auditory or audiovisual integration regions (Dehaene et al., 2010; Desroches et al., 2010; Yoncheva, Zevin, Maurer, & McCandliss, 2010).

The suppressive interaction based on the additive model [AV - (A + V)] was found to be significant in multiple temporal and parietal regions, which partly overlaps with the findings of audiovisual integration in the superior temporal cortex in adults (Blau et al., 2008; Raij et al., 2000; van Atteveldt et al., 2004). The dorsal (temporoparietal) cortical network, including supramarginal gyrus and angular gyrus in the inferior parietal cortex and the posterior superior temporal gyrus (pSTG) is related to mapping print onto its phonological and semantic representation (Sandak, Mencl, Frost, & Pugh, 2004). The results from Study II suggested that a more widely distributed temporoparietal cortical network is recruited to support learning the association of orthography with phonological codes in beginner readers (Pugh et al., 2013). In addition, the rather late time window of the suppressive integration indicated a less automatic audiovisual process in children (Blomert, 2011; Froyen et al., 2009), which might involve top-down modulation.

Suppressive integration in temporoparietal regions was correlated with phonological processing, rapid automatized naming as well as reading and writing skills after controlling the age effect. More specifically, phonological skills were correlated with the interaction effect in the right precuneus and inferior parietal regions, while rapid naming of letters was correlated with the interaction effect in the left supramarginal and right precuneus. Similar results have been reported in previous studies on the associations between pre-reading skills (phonological processing and rapid naming) and brain changes in temporoparietal regions in children (Raschle, Chang, & Gaab, 2011; Raschle, Zuk, & Gaab, 2012; Specht et al., 2009). Moreover, the integration effect in the right precuneus was consistently correlated with reading skills, for example, nonword list and nonword text and word list reading accuracy. This was in consonance with one study (Pugh et al., 2013), which also used the brain (fMRI) and behavior correlation approach and reported that the activation in the precuneus to prints and speech sounds (pseudowords/words) is correlated with reading-related skills.

The congruency effect (AVC vs. AVI) was not significant in Study II. Since many of the participants in this study had only one or two years of reading instruction, it is likely that at the neural level, they may not establish fully automated letter–speech sound integration, as shown by earlier studies using MMN (Blomert, 2011; Froyen et al., 2009). In addition, the audiovisual congruency effect is not only heavily dependent on the experimental task (Andersen et al., 2004; van Atteveldt, Formisano, Goebel, & Blomert, 2007) but also seemed to interact with the imaging methods. For example, fMRI studies (Blau et al., 2010; Brem et al., 2010) on children have found congruency effects using similar implicit active task, but not during explicit matching task (van Atteveldt, Formisano, Goebel, & Blomert, 2007), whereas an active matching task was able to elicit a congruency effect using MEG (Raij et al., 2000). Taken together, the suppressive integration and congruency effects in Study II indicated more general audiovisual integration processes in children who have not obtained the fully automatic level of letter–speech sound integration, as evidenced by the absence of congruency effect.

According to the general neurodevelopmental model of reading proposed by (Cornelissen, Hansen, Kringelbach, & Pugh, 2010; Pugh et al., 2001), the temporoparietal (dorsal) brain networks are crucial during the early phase of learning to read. Working together with the anterior circuits (especially the inferior frontal region), the dorsal (temporoparietal) reading system is involved in the emergence of phonological awareness (Katzir, Misra, & Poldrack, 2005) and in integration between orthography, phonology, and semantics (Pugh et al., 2001). The maturation of the dorsal reading circuit will then guide and support the development of the left ventral (occipitotemporal) circuit, including the VWFA (Dehaene & Cohen, 2011) for supporting fluent reading in advanced readers. Therefore, Study II underscores the crucial role of the temporoparietal circuits in developing phonological awareness and initiating automatic letter–speech sound associations in beginning readers.

4.3 Audiovisual learning in the human brain

Study III investigated the neural mechanisms of letter–speech sound association learning. The brain dynamics during initial learning and memory consolidation after learning were captured in a two-day letter–speech sound learning experiment using magnetoencephalography. The MEG experiment was designed to separate the audiovisual processing and the grapheme–phoneme associative learning by consecutive presentations of, first, audiovisual stimuli, and second, different learning cues. Two sets of audiovisual stimuli were used for training in which the letter–speech sound association in one set (Learnable) could be learned but not in the other set (Control) based on the different learning cues provided. The participants' performance was monitored with trial-by-trial precision in the testing blocks after each learning block. Changes related to associative learning were examined by comparing the Learnable and Control conditions at different

learning stages. Dynamic brain changes were found during multisensory learning and, most interestingly, during the processing of the learning cues.

The brain responses to the Learnable and Control visual stimuli (presented alone) showed rather stable differences after the learning of audiovisual associations. More specifically, Learnable and Control letters started to elicit different activation patterns around left temporoparietal, paracentral and occipital regions at late learning stages on Day 1 (learning index: >4) and also on Day 2. The Learnable letters were linked to their phonemic representations through training, and these audiovisual connections were strengthened over time. This could lead to a different processing mechanism, in which only the level of orthographic familiarity was increased for the Control stimuli. Similar results have been reported by earlier studies comparing single letters and pseudo-letters (Bann & Herdman, 2016; Herdman & Takai, 2013). The relatively late time windows (455 ms on Day 1 and 380 ms on Day 2) of the significant cluster might reflect the still slower processing speed of newly-learned grapheme-phoneme mapping than well-established or over-learned ones (Brem et al., 2018; Herdman & Takai, 2013; Maurer et al., 2005). The location of significant difference in temporoparietal also matches with previous findings on early reading acquisitions (Carreiras, Quiñones, Hernández-Cabrera, & Duñabeitia, 2015; Dehaene, Cohen, Morais, & Kolinsky, 2015; Pugh et al., 2001) and artificial word training (Quinn et al., 2017), which showed that these dorsal circuits are important for grapheme to phoneme conversion (Pugh, Mencl, Jenner, et al., 2000; Sandak et al., 2004; Taylor et al., 2014). However, no difference was found for the auditory only conditions between Learnable and Control sets, which suggested that mapping of additional visual letters to existing phonemes might not alter the brain representations of the existing phonemes (Familiar Finnish phonemes) itself, but rather a new audiovisual association was formed.

Region-of-interest (ROI) (left and right pSTS) analysis of audiovisual congruency effects based on the ANOVA model showed no difference before the learning (learning index: 0) as expected. Brain responses to the Learnable (AVC and AVI) and Control (AVX) sets started to differ in the early learning stage on Day 1 (learning index:1-4), which indicated that categorization of Learnable and Control sets seemed to be easier and appear earlier than the learning of the audiovisual association in the Learnable set. It was only at a later stage (learning index: >4) in the testing blocks on Day 1 when the congruency (AVC > AVI) effect, a brain-level index of learned associations was found to be significant. This was consistent with earlier studies using similar grapheme-phoneme training paradigms (Karipidis et al., 2018; Karipidis et al., 2017) as well as for the over-learned audiovisual stimuli (Raij et al., 2000; van Atteveldt et al., 2004). Moreover, this congruency (AVC > AVI) effect in the testing blocks was absent on Day 2 and only responses to Learnable audiovisual congruent (AVC) and Control (AVX) stimuli showed a significant difference in the training blocks on Day 2. These changes highlighted the dynamic characteristics of brain processes related to the newly-learned audiovisual associations. Memory consolidation and

reorganization during overnight sleep seems to affect multisensory processing at the initial stage of audiovisual learning.

Converging evidence for the dynamic audiovisual processing in the early learning stage was also found in audiovisual suppressive interaction effects based on the additive model (A + V vs. AV). The suppressive effect (A + V > AV) showed differences between Learnable and Control conditions only at the early learning stage (learning index:1-4) on Day 1 in the left parietal region which had been indicated to be crucial for grapheme-phoneme mapping in early reading acquisition (Pugh et al., 2013; Sandak et al., 2004). As discussed in Study I and Study II, the suppressive interaction effect reflects a more general form of cross-modal interaction and could be engaged transiently during the early learning stage (as shown by Study II) before a stable integration of the two modalities was established.

Overall, the above discussion on uni/multi-modal processing of learned associations suggested that the audiovisual processing is very dynamic and depends on the different learning stages and tasks, whereas the brain representation of learned letters seemed to be more stable and persistent after successful learning of its phonological association. These early dynamic processes have not been reported before since most earlier studies had only examined the multisensory or learning effects at one time point after training.

The brain responses to the three different learning cues provided a unique window into the brain mechanism of associative learning. In general, a reversed pattern was observed for the learning cue processing as compared to the audiovisual processing: the brain responses to the learning cues were different mainly before and immediately after (learning index: 0-4) behavioral learning could be observed on Day 1 and no difference was observed between the brain activations to three learning cues on Day 2. The audiovisual associative learning (\surd vs. X contrast) and non-associative learning (\boxtimes vs. X contrast) effects showed largely overlapping in brain regions around the left and right middle and inferior temporal and some deeper brain sources near the insula and bilateral medial temporal (hippocampus) regions. Similar results have been found in previous studies, for example, the inferior temporal cortex has been indicated to be crucial in forming cross-modal associations (Gibson & Maunsell, 1997; Miyashita & Hayashi, 2000; Sakai & Miyashita, 1991) and the hippocampus and nearby areas are related to working memory processes (Olson, Moore, Stark, & Chatterjee, 2006; Quak, London, & Talsma, 2015; Yonelinas, 2013). The audiovisual associative and non-associative learning (\surd vs. \boxtimes contrast) showed different activations in parts of the left temporal region and right insula. The decreased activation in the left temporal cortex for audiovisual associative learning compared with non-associative learning might be related to the cross-modal memory encoding (Tanabe, Honda, & Sadato, 2005), and the increased activation in the right insula regions might be related to multisensory attention (Chen et al., 2015). Such subtle differential activation strength between audiovisual associative and non-associative learning in both hemispheres, which probably reflects the unique cognitive processes in associative learning.

Learning speed in artificial grapheme–phoneme association training has been shown to correlate with future reading fluency and was suggested as a novel tool to identify children with future reading problems (Karipidis et al., 2018; Karipidis et al., 2017). Audiovisual non-associative learning speed in Study III showed a correlation with rapid naming ability, which is a robust behavioral precursor of reading fluency across various languages (Kirby et al., 2010; Moll et al., 2014). Rapid naming has been found to be linked to cross-modal learning by other artificial learning studies (Aravena, Tijms, Snellings, & van der Molen, 2018; Karipidis et al., 2018; Karipidis et al., 2017). In our case, the correlation with associative learning was not significant after FDR correction could be due to statistical power issues ($r = -0.45$ and uncorrected $p = 0.012$ for the correlation between RAN objects and the associated learning speed). It could also be possible that experiment design seemed to favor the learning of separating the Learnable and Control stimuli from each other in addition to the learning of audiovisual associations.

4.4 General discussion

This dissertation investigated various aspects of brain changes related to learning audiovisual associations in reading acquisition using magnetoencephalography. Study I extended the previous findings on well-learned audiovisual integration of alphabetic languages to logographic languages, which include the semantic process for each character–speech sound pair. In Study II, audiovisual integration was examined in beginning readers, and brain–behavior analyses were utilized to examine the relationship between children’s cognitive skills and cortical responses related to auditory, visual, and audiovisual processes of letters and speech sounds. Study III was designed to capture the brain dynamics of learning grapheme–phoneme associations using a well-controlled audiovisual training paradigm.

In general, across all three studies, the audiovisual processing showed large similarities in terms of time windows and brain locations. For example, the superior temporal region, which has been shown to be important for audiovisual integration of letter–speech sound (van Atteveldt et al., 2004) and audiovisual objects (Beauchamp, Lee, et al., 2004) in general, was also identified to be an important cortical hub for audiovisual integration in this dissertation work. In Study I, the left superior temporal cortex was actively involved in the processing of both multisensory interaction ([A + V vs. AV]) and congruency comparison. This showed that the STC was indeed a common region of audiovisual processing across languages, including both logographic and alphabetic languages. In Study II, the suppressive interaction was also identified in various temporoparietal regions, including the superior regions for the integration of letter–speech sounds in children learning to read. In Study III, the activity in superior temporal areas was found to be sensitive to the effect of learning audiovisual associations and seemed to be dynamic across different learning

stages. In addition, the time windows of audiovisual integration across three studies were also similar (after about 400 ms). This relatively late time window supports the idea that audiovisual integration was only possible after the unisensory processing of auditory and visual information.

The audiovisual processing also showed some distinct differences across orthographies and more important at different stages of learning to read. In Study I, different suppressive interaction effects found in Chinese and Finnish groups indicated the adaptive nature of the cortical networks of the audiovisual processing. Importantly, in addition to the integration in STC, the inferior frontal cortex was also involved in audiovisual processing of semantic information in native Chinese speakers. The findings in the left superior temporal cortex are surprisingly similar to those in alphabetic languages but with also some unique brain processes in the inferior frontal cortex specific to the logographic scripts. The results from Study I thus suggested that there are some universal audiovisual integration mechanisms in reading acquisition complemented by additional language-specific processes. Study II showed that at the early stages of learning the letter–speech sound integration, more broad brain regions (left and right temporoparietal regions) seemed to be recruited for the audiovisual processing. Findings from Study II point to the important role of these temporoparietal regions in learning letter–speech sound associations in early reading development. Study III further confirmed the dynamic nature of audiovisual processing using a cross-modal association learning task in two days. Depending on the different stages of learning grapheme–phoneme associations, rapid changes in brain activity have been observed in a broad range of temporoparietal brain regions. This might be related to activations of brain circuits for forming multisensory memory and subsequent overnight consolidation. Study III further indicated that the newly learned cross-modal association also affects the visual representation of letters in distributed occipital and parietal regions.

Based on findings from this dissertation and previous studies, a model regarding the learning of audiovisual associations in reading acquisition was proposed in Figure 4. In this model, the auditory (e.g., the sound of /a/) and visual (e.g., letter *a*) sensory inputs are first processed in the primary auditory and visual cortices. The auditory features of the stimuli are then combined to form more abstract representations most likely in the superior temporal regions in both early and late time windows as reflected for example by the auditory P2 response (Hämäläinen et al., 2019) and late sustained responses (Ceponiene, Alku, Westerfield, Toriki, & Townsend, 2005). The visual features have been processed, and they form an abstract representation when the visual information is processed along the vOT cortex known to respond to orthographic regularities. The auditory and visual information is then integrated in the multisensory areas in the superior temporal cortex (marked in red in Figure 4) (Beauchamp, Lee, et al., 2004; Raji et al., 2000; van Atteveldt et al., 2009) to form a coherent audiovisual object at a relatively late time window after the auditory and visual inputs are processed (see van Atteveldt et al., 2009 for a functional neuro-anatomical model of letter–speech sound integration in literate adults). Depending on the different

tasks used during the experiment, there might be top-down modulatory feedback from the frontal regions (marked in yellow in Figure 4) to the multisensory regions (van Atteveldt, Formisano, Goebel, & Blomert, 2007). In addition, Study I further pointed out that semantic-related processing involved the left inferior frontal regions.

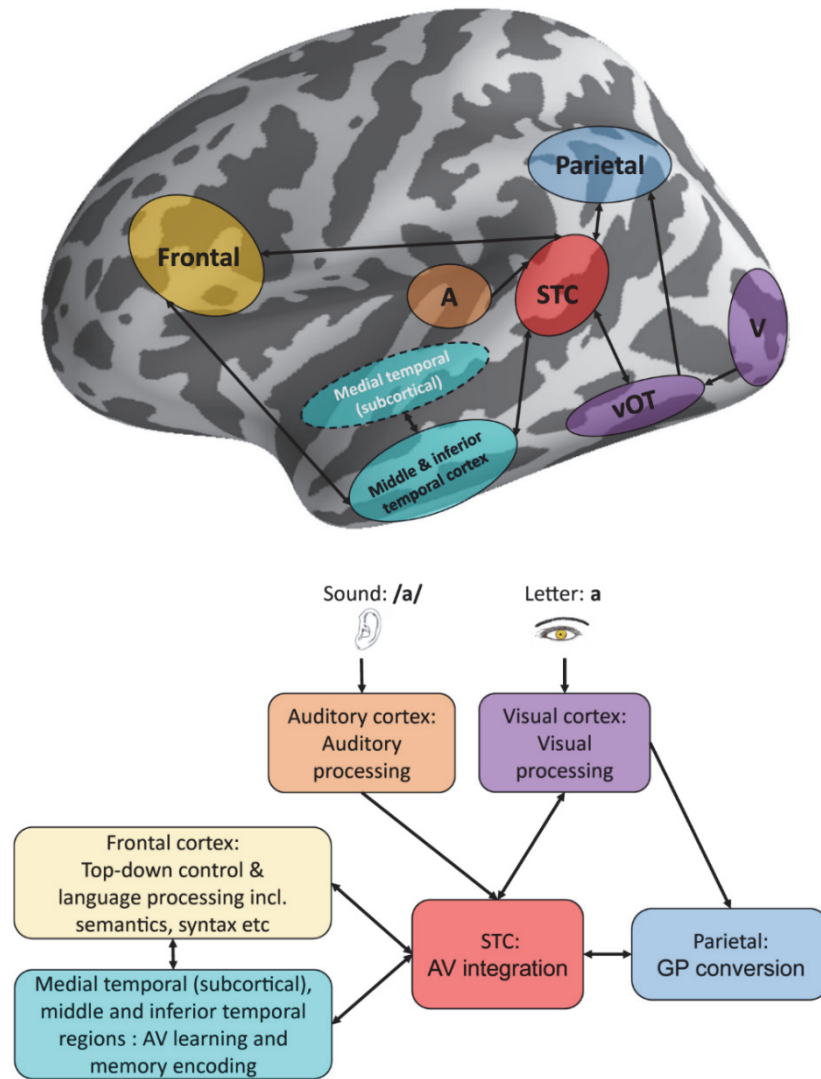


FIGURE 4 Schematic diagram of the possible network involved in the learning of letter-speech sound associations. A = Auditory cortex, V = Visual cortex, STC = superior temporal cortex, vOT = ventral occipitotemporal cortex, GP = Grapheme-phoneme.

During the initial learning stage, the audiovisual representations are encoded, and short-term memory of the audiovisual objects are stored in the middle and inferior temporal and possibly also in the medial temporal (e.g., hippocampus) regions (marked in cyan in Figure 4) (Easton & Gaffan, 2001; Quinn et al., 2017). Frontal regions (Mei et al., 2014) have been suggested to be involved in the top-down control mechanism, for example, in the selection of

cross-modal features (Calvert et al., 2001; Hämäläinen et al., 2019) to combine and direct attention to the relevant learning cues. In addition, parietal regions (marked in blue in Figure 4) also receive visual inputs (of letters) from the occipital regions and might be involved in storing the corresponding phonological representation of the letters (grapheme-phoneme conversion) by interacting with the multisensory superior temporal cortex during the early stages of learning as indicated in Study III. As learning progress, changes have been reported to occur in vOT (Brem et al., 2010; Brem et al., 2018; Hashimoto & Sakai, 2004; Madec et al., 2016; Quinn et al., 2017) and dorsal pathway (parietal regions) (Hashimoto & Sakai, 2004; Mei et al., 2015; Mei et al., 2014; Taylor et al., 2017; Taylor et al., 2014) as well as the STC (Hämäläinen et al., 2019; Karipidis et al., 2018; Karipidis et al., 2017; Madec et al., 2016) for forming optimal cortical representation and automatic processing of the audiovisual objects.

A letter-speech sound learning deficit has been reported as one possible key factor for dyslexia in studies using artificial letter training paradigms (Aravena, Snellings, Tijms, & van der Molen, 2013; Aravena et al., 2018; Karipidis et al., 2018; Karipidis et al., 2017). Findings from this dissertation could provide a better understanding of neural dynamics that underpin grapheme-phoneme learning and could be used to identify specific bottlenecks in learning cross-modal associations in children. A more refined grapheme-phoneme learning model could provide better scientific evidence on how to improve the teaching of multimodal material. For example, teachers could adapt individual training programs that are targeted to strengthen the specific pathways in the model for enhancing children's learning efficiency.

4.5 Limitations

The task during the audiovisual experiment could affect the specific audiovisual processing observed as shown in earlier studies (van Atteveldt et al., 2009; van Atteveldt, Formisano, Goebel, & Blomert, 2007). In Study I and Study II, active tasks were used for investigating the audiovisual integration of learned associations. Active audiovisual paradigm is more realistic since in real-life situations the audiovisual processing is an active action (e.g., learning to read). However, it also brought certain complications in analyzing the data and interpreting results. For example, Study I used a dual-modality working memory task, in which the working memory processes might interfere with audiovisual integration. In addition, since the stimuli consisted of Chinese characters and speech sounds, the working memory task might have different demands for native Chinese speakers and Finnish speakers who have never learned Chinese before. Therefore, direct group comparison was not conducted considering that the task demand and the underlying brain mechanism of audiovisual integration might be quite different for Chinese and Finnish groups.

The accuracy of MEG source localization could potentially restrict the conclusions on the precise brain areas. In Study I and Study III, the FsAverage

brain template was used for MEG source localization since individual MRI images were not acquired due to limited resources. To minimize the mismatch of brain size and shape between the individual brain and the template brain, a three-parameter scaling was used to fit the template brain to the individual digitized head points. In general, compared with other functional neuroimaging techniques such as fMRI, MEG is not the optimal tool for accurate localization of brain activity. However, compared with EEG, MEG is insensitive to the head tissue conductivity, which would allow better brain source estimations. The main goal of the present research was not to localize the brain activities with millimeter accuracy but to have a rough estimation of brain activities over relatively large cortical surfaces. Therefore MEG could provide relatively good localization accuracy even with brain templates and has the advantage of high temporal resolution (millisecond). In Study II, although the size of the child's brain was smaller than adults and thus was further away from the MEG sensors, individual structural MRIs were acquired to improve the localization accuracy. Deep brain structures such as the medial temporal system (including the hippocampus) play a crucial role in the learning and memory processes, as reported in a number of studies (Axmacher, Elger, & Fell, 2008; Brasted, Bussey, Murray, & Wise, 2003; Jarrard, 1993; Mayes, Montaldi, & Migo, 2007). MEG is not optimal to localize deep brain activity due to the decreased signal-to-noise ratio (SNR) as a function of source depth. However, evidence suggests that hippocampal activities could be captured with MEG (Attal & Schwartz, 2013; Ruzich, Crespo-García, Dalal, & Schneiderman, 2019), especially during learning and memory tasks (Backus, Schoffelen, Szabéni, Hanslmayr, & Doeller, 2016; Shah-Basak et al., 2018; Taylor, Donner, & Pang, 2012). In Study III, some activities related to the processing of the learning cues seemed to emerge from the deep brain sources. However, due to the limited SNR and spatial resolution in MEG, caution should be taken when interpreting results, particularly the localization regarding the deep brain sources.

In Study III, the learning process was tracked on two consecutive days; ideally it would be interesting to track the learning process over a longer period (e.g., one week). Several other studies have looked at the behavioral (Aravena et al., 2013; Aravena et al., 2018) and neural (Taylor et al., 2017; Taylor, Davis, & Rastle, 2019) change related to cross-modal learning over an extended period of time. However, the main goal of Study III was to investigate the brain mechanisms during the initial learning using an artificial grapheme-phoneme training paradigm simulating the situation when children typically learn letters. Interesting brain dynamics were observed during letter-speech sound learning in Study III, which could provide new information on potential brain mechanisms leading to long-term learning outcomes. The grapheme-phoneme learning was examined in adults in Study III based on the assumption that this kind of basic multisensory learning is preserved in adults and would recruit similar brain networks as in children learning to read. Therefore a potential interesting research question concerns the difference of brain mechanisms in adults and children when learning novel letter-speech sound associations.

4.6 Future directions

A more detailed neurocognitive model, including functional connectivity and network pattern, is needed for a better understanding of the cortical organization and the developmental trajectory of letter–speech sound learning. Learning and integrating cross-modal associations involves interaction and communication between multiple brain regions, including the sensory processing regions, the short-term memory systems, attention and cognitive control systems in frontal and parietal areas. Therefore functional brain connectivity and network analysis would be an interesting approach for future studies on letter–speech sound association integration and learning. The current dissertation work has identified important brain regions (hubs) and time windows that could be useful for future studies that use optimal experimental design for connectivity and network analysis.

The automatic letter–speech sound integration deficit has been proposed as an important cause of reading difficulties (Blomert, 2011; Froyen et al., 2011; Žarić et al., 2014). Several studies (Aravena et al., 2018; Karipidis et al., 2017; Pleisch et al., 2019) have tried to examine the difference of LSS integration at different levels of learning to read in children with reading difficulties and typically developing controls. However, less is known about the differences in brain processes during the learning of letter–speech sound associations. For that reason future studies could apply the developed LSS learning paradigm (similar to the learning experiment implemented in Study III) to examine the formation of letter–speech sound associations in children with varying reading skills, including dyslexic readers. Such studies could help in understanding the neural mechanisms of reading difficulties and may inspire the design of tools (e.g., dynamic assessment) for early identification of such problems (Aravena et al., 2018; Elbro, Daugaard, & Gellert, 2012). Furthermore, the brain-behavior analysis could be utilized to identify brain indexes of audiovisual learning, which correlated with children’s cognitive skills. Since the artificial grapheme–phoneme training approach is not restricted to a specific language, it could provide a more general and flexible measure of children’s cross-modal learning ability even before formal reading instruction. This approach could potentially benefit the early identification (Elbro et al., 2012) and intervention of children with future reading problems and provide targeted personalized training for those with reading difficulties.

Successful letter–speech sound integration is a crucial step for fluent reading, yet reading is a much more complex cognitive process that includes, for example, syntactic, semantic, and phonological processes to extract complete information from the text in close temporal and spatial proximity. Linguistic elements at one level are combined to construct linguistic elements at a higher level; for example, letters are combined into words that contain meanings as well as sounds. In alphabetic languages, how the letter–speech sound integration interacts with high-level linguistic processes (e.g., at the word or even sentence level) during learning to read is still poorly understood. However, due to the

complexity of natural languages, it is immensely difficult to isolate the exact brain mechanism responsible for specific linguistic processes and interactions at different levels. Artificial language training paradigms (Folia, Uddén, De Vries, Forkstam, & Petersson, 2010) provide a simplified linguistic structure and good control for the prior learning experience, and they are useful tools for investigating brain mechanisms of learning to read from the basic letter-speech sound processing (Study III) to higher-level linguistic processes and the interaction between different levels.

In conclusion, the results of this dissertation indicate the adaptive nature of the audiovisual process. It was found that at the initial learning stage, a dynamic and distributed cortical network was recruited during the forming and consolidating of cross-modal associations. The cortical audiovisual process was confirmed in this dissertation to still be at the immature state in children learning to read, but the level of automaticity was associated with their reading-related skills. In addition, audiovisual integration in reading seems to recruit certain universal audiovisual integration mechanisms and is complemented by language-specific processes.

YHTEENVETO (SUMMARY)

Audiovisuaalisten yhteyksien oppimiseen liittyvät aivotoiminnan muutokset lukemaan oppimisen yhteydessä

Kirjainten ja äänteiden yhteyden oppiminen on aakkosia käyttävissä kielissä yksi ensimmäisiä askeleita kohti lukutaidon omaksumista. Tutkimukset ovat osoittaneet, että neuraalisella tasolla tämä audiovisuaalinen integraatioprosessi näyttää olevan täysin automaattinen vasta, kun lukemista on harjoiteltu useita vuosia ja että kykenemättömyys muodostaa automaattisesti grafeemi-foneemi-vastavuuksia on yksi lukemisvaikeuksien taustalla olevista syistä. Olemassa oleva audiovisuaalisen integraation neuroanatominen malli perustuu pääasiassa tutkimustuloksiin, jotka koskevat lukutaitoisten, aakkoskirjoitusta käyttävien aikuisten jo opittua kirjainten ja äänteiden yhdistämistä. Lukemaan opettelun alkuvaiheessa tapahtuvasta audiovisuaalisesta prosessoinnista tiedetään kuitenkin vähemmän, samoin kuin audiovisuaalisesta integraatiosta esimerkiksi logografista eli sanakirjoitusta käyttävissä kirjoitusjärjestelmissä.

Tämän väitöstutkimuksen tavoitteena oli selvittää lukemisen audiovisuaalisten yhteyksien oppimiseen liittyviä aivotoiminnan muutoksia. Erityisesti keskityttiin aivodynamiikan muutoksiin kirjain-ääne-yhdistelmien oppimisen aikana (tutkimus 3) ja pian niiden oppimisen jälkeen (tutkimus 2). Lisäksi tutkimuksessa 1 kartoitettiin audiovisuaalista prosessointia logografista kirjoitusjärjestelmää käyttävässä kielessä (kiina), jossa kullakin kirjoitusmerkillä on oma merkitys. Magnetoenkefalografia (MEG) on kuvantamistekniikka, jolla aivotointaa voidaan kartoittaa aivojen sähköisen toiminnan tuottamia magneettikenttiä taltioimalla. Sitä käytettiin kaikissa kolmessa osatutkimuksessa, jotta voitiin tallentaa yksiaistisen ja moniaistisen prosessoinnin tuottamaa aivotointaa.

Tutkimuksessa 1 selvitettiin audiovisuaalista integraatioprosessia kiinan kielessä, jonka merkit liittyvät tavun äänneeseen ja merkitykseen. Syntyperäisten kiinan puhujien ryhmä ja toinen vastaava suomenkielinen kontrolliryhmä osallistuivat audiovisuaaliseen MEG-kokeeseen, jossa ärsykkeinä käytettiin kiinan kirjoitusmerkkejä ja puheäänteitä. Suomalaisessa ryhmässä havaittiin vain supressioefekti oikeassa päälaki- ja takaraivolohkossa, mikä todennäköisesti liittyy yleiseen tuntemattomien audiovisuaalisten ärsykkeiden aiheuttamaan audiovisuaaliseen prosessiin. Kiinalaisessa ryhmässä havaittiin sekä audiovisuaalista supressio- että kongruenssiefektiä vasemman ylemmän ohimolohkon ja vasemman alemman otsalohkon alueilla. Tämä osoitti, että logografisessa kielessä audiovisuaalisessa prosessoinnissa käytettiin ylemmillä ohimolohkoalueilla samanlaista aivomekanismia kuin aakkosia käyttävissä kielissä, mutta mukaan valjastettiin lisäksi alemmat otsalohkoalueet sanojen merkitysten käsittelyyn.

Tutkimuksessa 2 keskityttiin lukemaan opettelevan suomalaisen lapsiryhmän aivotointaan yksiaistisen (auditiivisen/visuaalisen) ja audiovisuaalisen prosessoinnin aikana. Tarkoituksena oli selvittää yksiaististen ja audiovisuaalisten vasteiden yhteys lasten lukemiseen liittyviin kognitiivisiin taitoihin. Ryhmä

tyypillisesti kehittyviä 6–11-vuotiaita suomalaislapsia osallistui erityisesti lapsille suunniteltuun audiovisuaaliseen kokeeseen, jossa ärsykkeinä käytettiin suomen kirjaimia ja äänneitä. Aivoaktivaatio reaktiona auditiivisiin, visuaalisiin ja audiovisuaalisiin ärsykkeisiin, samoin kuin audiovisuaalista integraatiota mittaava aivovaste, korreloivat lukutaidon ja kognitiivisten taitojen kanssa, jotka ennustavat lukutaidon kehitystä. Regressioanalyysin perusteella myöhäinen auditiivinen vaste, noin 400 ms ärsykkeen alusta, osoitti selkeimmin yhteyttä fonologiseen prosessointiin ja nopeaan automaattiseen nimeämiskykyyn. Audiovisuaalinen integraatioefekti oli lisäksi huomattavin vasemmilla ja oikeilla temporoparietaalisilla alueilla, ja useilla näistä alueista aktivaatio oli yhteydessä lasten luku- ja kirjoitustaitoihin. Nämä tulokset viittasivat lasten audiovisuaalisen prosessoinnin kypsyttämiseen ja temporoparietaalisten alueiden merkitykseen lukemaan opetteluun alkuvaiheessa sekä niiden ainutlaatuisen rooliin lukutaidon omaksumisessa.

Tutkimuksessa 3 pyrittiin kartoittamaan niitä aivokuoren mekanismeja, jotka tukevat kirjainten ja äänneiden oppimista, sekä erityisesti aivojen dynamiikkaa grafeemi-foneemi-vastaavuuksia opittaessa. Yhteensä 30 suomalaista aikuista osallistui kahtena peräkkäisenä päivänä MEG-kokeeseen, jossa he opettelivat yhdistämään itselleen uusia vieraskielisiä kirjaimia tuttuihin suomalaisiin äänneisiin. Harjoituksessa käytettiin kahta audiovisuaalista ärsykejoukkoa: ensimmäisessä joukossa ("opittava") audiovisuaalinen yhteys voitiin oppia annettujen vihjeiden perusteella, toisessa joukossa ("kontrolli") sen sijaan ei. Oppimisen edistymistä seurattiin ärsyke kerrallaan luokitellen sen perusteella eri oppimisvaiheita. Aivojen heräteasteissa havaittiin dynaamisia muutoksia moniaistiseen prosessointiin sekä visuaaliseen kirjainten prosessointiin liittyen grafeemi-foneemi-vastaavuuksia opeteltaessa ja muistijäljen vahvistuttua yön aikana. Ylipäänsä aistien välinen oppimisprosessi näyttöisi säätelevän toimintaa laajojen aivoalueiden muodostamassa verkostossa, johon kuuluvat ylempi ohimolohko ja dorsaalinen hermorata. Yksi keskeinen löydös oli, että keski- ja alaohimolohkon alueet olivat mukana moniaistisessa muistien koodauksessa oppimisvihjeen käsittelyn aikana. Kolmannen osatutkimuksen tulokset korostavat kirjainten ja puheäänneiden vastaavuuksien oppimiseen liittyvää aivojen dynaamisuutta ja plasztisuutta sekä tarjoavat tarkemman grafeemi-foneemi-oppimisen mallin.

Kaiken kaikkiaan väitöstutkimuksen tulokset osoittivat, että audiovisuaalinen prosessointi mukautuu tehtävävaatimusten ja oppimisen seurauksena. Oppimisen alkuvaiheessa aistien välisiä yhteyksiä muodostettaessa ja vahvistettaessa käyttöön otettiin dynaaminen ja hajautettu aivokuoren verkosto. Aivokuorella tapahtuvan audiovisuaalisen prosessoinnin todettiin olevan lukemista opettelevilla lapsilla vielä kypsyttämättömän, mutta automaattisuuden taso oli yhteydessä heidän lukemiseen liittyviin taitoihinsa. Lisäksi luettaessa tapahtuva audiovisuaalinen integraatio näyttää hyödyntävän tiettyjä universaaleja audiovisuaalisia integraatiomekanismeja, joita täydennetään kielikohtaisilla prosesseilla.

REFERENCES

- Ahveninen, J., Jääskeläinen, I. P., Raij, T., Bonmassar, G., Devore, S., Hämäläinen, M., ... & Witzel, T. (2006). Task-modulated “what” and “where” pathways in human auditory cortex. *Proceedings of the National Academy of Sciences*, 103(39), 14608-14613.
- Andersen, T. S., Tiippana, K., & Sams, M. (2004). Factors influencing audiovisual fission and fusion illusions. *Cognitive Brain Research*, 21(3), 301-308.
- Aravena, S., Snellings, P., Tijms, J., & van der Molen, M. W. (2013). A lab-controlled simulation of a letter–speech sound binding deficit in dyslexia. *Journal of Experimental Child Psychology*, 115(4), 691–707.
- Aravena, S., Tijms, J., Snellings, P., & van der Molen, M. W. (2018). Predicting Individual Differences in Reading and Spelling Skill With Artificial Script-Based Letter–Speech Sound Training. *Journal of Learning Disabilities*, 51(6), 552–564.
- Attal, Y., & Schwartz, D. (2013). Assessment of subcortical source localization using deep brain activity imaging model with minimum norm operators: a MEG study. *PLoS One*, 8(3), e59856.
- Axmacher, N., Elger, C. E., & Fell, J. (2008). Ripples in the medial temporal lobe are relevant for human memory consolidation. *Brain*, 131(7), 1806-1817.
- Backus, A. R., Schoffelen, J. M., Szabéniyi, S., Hanslmayr, S., & Doeller, C. F. (2016). Hippocampal-prefrontal theta oscillations support memory integration. *Current Biology*, 26(4), 450-457.
- Bann, S. A., & Herdman, A. T. (2016). Event Related Potentials Reveal Early Phonological and Orthographic Processing of Single Letters in Letter-Detection and Letter-Rhyme Paradigms. *Frontiers in Human Neuroscience*, 10, 176.
- Beauchamp, M. S., Argall, B. D., Bodurka, J., Duyn, J. H., & Martin, A. (2004). Unraveling multisensory integration: patchy organization within human STS multisensory cortex. *Nature Neuroscience*, 7(11), 1190–1192.
- Beauchamp, M. S., Lee, K. E., Argall, B. D., & Martin, A. (2004). Integration of auditory and visual information about objects in superior temporal sulcus. *Neuron*, 41(5), 809–823.
- Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal statistical society: series B (Methodological)*, 57(1), 289-300.
- Bernstein, L. E., Auer, E. T., & Takayanagi, S. (2004). Auditory speech detection in noise enhanced by lipreading. *Speech Communication*, 44(1), 5–18.
- Besle, J., Fort, A., & Giard, M. H. (2004). Interest and validity of the additive model in electrophysiological studies of multisensory interactions. *Cognitive Processing*, 5(3), 189-192.
- Blau, V., Reithler, J., van Atteveldt, N., Seitz, J., Gerretsen, P., Goebel, R., & Blomert, L. (2010). Deviant processing of letters and speech sounds as

- proximate cause of reading failure: a functional magnetic resonance imaging study of dyslexic children. *Brain*, 133(3), 868-879.
- Blau, V., van Atteveldt, N., Ekkebus, M., Goebel, R., & Blomert, L. (2009). Reduced neural integration of letters and speech sounds links phonological and reading deficits in adult dyslexia. *Current Biology*, 19(6), 503-508.
- Blau, V., van Atteveldt, N., Formisano, E., Goebel, R., & Blomert, L. (2008). Task - irrelevant visual letters interact with the processing of speech sounds in heteromodal and unimodal cortex. *European journal of neuroscience*, 28(3), 500-509.
- Blomert, L. (2011). The neural signature of orthographic-phonological binding in successful and failing reading development. *NeuroImage*, 57(3), 695-703.
- Bönstrup, M., Iturrate, I., Thompson, R., Cruciani, G., Censor, N., & Cohen, L. G. (2019). A rapid form of offline consolidation in skill learning. *Current Biology*, 29(8), 1346-1351.
- Bonte, M., & Blomert, L. (2004). Developmental changes in ERP correlates of spoken word recognition during early school years: a phonological priming study. *Clinical Neurophysiology*, 115(2), 409-423.
- Bonte, M., Correia, J. M., Keetels, M., Vroomen, J., & Formisano, E. (2017). Reading-induced shifts of perceptual speech representations in auditory cortex. *Scientific Reports*, 7(1), 5143.
- Brasted, P. J., Bussey, T. J., Murray, E. A., & Wise, S. P. (2003). Role of the hippocampal system in associative learning beyond the spatial domain. *Brain*, 126(5), 1202-1223.
- Bregman, A. S. (1994). *Auditory Scene Analysis: The Perceptual Organization of Sound*. MIT Press.
- Brem, S., Bach, S., Kucian, K., Kujala, J. V., Guttorm, T. K., Martin, E., ... & Richardson, U. (2010). Brain sensitivity to print emerges when children learn letter-speech sound correspondences. *Proceedings of the National Academy of Sciences*, 107(17), 7939-7944.
- Brem, S., Hunkeler, E., Mächler, M., Kronschnabel, J., Karipidis, I. I., Pleisch, G., & Brandeis, D. (2018). Increasing expertise to a novel script modulates the visual N1 ERP in healthy adults. *International Journal of Behavioral Development*, 42(3), 333-341.
- Bremmer, F., Schlack, A., Shah, N. J., Zafiris, O., Kubischik, M., Hoffmann, K. P., ... & Fink, G. R. (2001). Polymodal motion processing in posterior parietal and premotor cortex: a human fMRI study strongly implies equivalencies between humans and monkeys. *Neuron*, 29(1), 287-296.
- Calvert, G. A. (2001). Crossmodal processing in the human brain: insights from functional neuroimaging studies. *Cerebral Cortex*, 11(12), 1110-1123.
- Calvert, G. A., Bullmore, E. T., Brammer, M. J., Campbell, R., Williams, S. C., McGuire, P. K., ... & David, A. S. (1997). Activation of auditory cortex during silent lipreading. *science*, 276(5312), 593-596.

- Calvert, G. A., & Campbell, R. (2003). Reading speech from still and moving faces: the neural substrates of visible speech. *Journal of Cognitive Neuroscience*, 15(1), 57–70.
- Calvert, G. A., Campbell, R., & Brammer, M. J. (2000). Evidence from functional magnetic resonance imaging of crossmodal binding in the human heteromodal cortex. *Current biology*, 10(11), 649–657.
- Calvert, G. A., Hansen, P. C., Iversen, S. D., & Brammer, M. J. (2001). Detection of audio-visual integration sites in humans by application of electrophysiological criteria to the BOLD effect. *NeuroImage*, 14(2), 427–438.
- Calvert, G. A., & Thesen, T. (2004). Multisensory integration: methodological approaches and emerging principles in the human brain. *Journal of Physiology-Paris*, 98(1-3), 191–205.
- Cappe, C., Thut, G., Romei, V., & Murray, M. M. (2010). Auditory–visual multisensory interactions in humans: timing, topography, directionality, and sources. *Journal of Neuroscience*, 30(38), 12572–12580.
- Carreiras, M., Quiñones, I., Hernández-Cabrera, J. A., & Duñabeitia, J. A. (2015). Orthographic Coding: Brain Activation for Letters, Symbols, and Digits. *Cerebral Cortex*, 25(12), 4748–4760.
- Ceponiene, R., Alku, P., Westerfield, M., Torki, M., & Townsend, J. (2005). ERPs differentiate syllable and nonphonetic sound processing in children and adults. *Psychophysiology*, 42(4), 391–406.
- Čeponienė, R., Shestakova, A., Balan, P., Alku, P., Ylaguchi, K., & Naatanen, R. (2001). Children's auditory event-related potentials index sound complexity and “speechness”. *International Journal of Neuroscience*, 109(3-4), 245–260.
- Čeponienė, R., Torki, M., Alku, P., Koyama, A., & Townsend, J. (2008). Event-related potentials reflect spectral differences in speech and non-speech stimuli in children and adults. *Clinical Neurophysiology*, 119(7), 1560–1577.
- Chen, T., Michels, L., Supekar, K., Kochalka, J., Ryali, S., & Menon, V. (2015). Role of the anterior insular cortex in integrative causal signaling during multisensory auditory--visual attention. *The European Journal of Neuroscience*, 41(2), 264–274.
- Cohen, L., Dehaene, S., Naccache, L., Lehéricy, S., Dehaene-Lambertz, G., Hénaff, M. A., & Michel, F. (2000). The visual word form area: spatial and temporal characterization of an initial stage of reading in normal subjects and posterior split-brain patients. *Brain*, 123(2), 291–307.
- Cohen, Y. E. (2009). Multimodal activity in the parietal cortex. *Hearing Research*, 258(1-2), 100–105.
- Cohen, Y. E., & Andersen, R. A. (2004). Multimodal Spatial Representations in the Primate Parietal Lobe. In C. Spence & J. Driver (Eds.), *Crossmodal Space and Crossmodal Attention* (pp. 154–176). Oxford University Press.
- Cornelissen, P., Hansen, P., Kringelbach, M., & Pugh, K. (2010). *The Neural Basis of Reading*. Oxford University Press.

- 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.
- Davis, M. H., Di Betta, A. M., Macdonald, M. J. E., & Gaskell, M. G. (2009). Learning and consolidation of novel spoken words. *Journal of Cognitive Neuroscience*, 21(4), 803-820.
- Dehaene, S., & Cohen, L. (2011). The unique role of the visual word form area in reading. *Trends in Cognitive Sciences*, 15(6), 254-262.
- Dehaene, S., Cohen, L., Morais, J., & Kolinsky, R. (2015). Illiterate to literate: behavioural and cerebral changes induced by reading acquisition. *Nature Reviews. Neuroscience*, 16(4), 234-244.
- Dehaene, S., Le Clec'H, G., Poline, J.-B., Le Bihan, D., & Cohen, L. (2002). The visual word form area: a prelexical representation of visual words in the fusiform gyrus. *Neuroreport*, 13(3), 321-325.
- Dehaene, S., Pegado, F., Braga, L. W., Ventura, P., Nunes Filho, G., Jobert, A., ... & Cohen, L. (2010). How learning to read changes the cortical networks for vision and language. *science*, 330(6009), 1359-1364.
- Denckla, M. B., & Rudel, R. G. (1976). Rapid "automatized" naming (RAN): Dyslexia differentiated from other learning disabilities. *Neuropsychologia*, 14(4), 471-479.
- Desikan, R. S., Ségonne, F., Fischl, B., Quinn, B. T., Dickerson, B. C., Blacker, D., ... & Albert, M. S. (2006). An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *Neuroimage*, 31(3), 968-980.
- Desroches, A. S., Cone, N. E., Bolger, D. J., Bitan, T., Burman, D. D., & Booth, J. R. (2010). Children with reading difficulties show differences in brain regions associated with orthographic processing during spoken language processing. *Brain Research*, 1356, 73-84.
- Diehl, M. M., & Romanski, L. M. (2014). Responses of prefrontal multisensory neurons to mismatching faces and vocalizations. *Journal of Neuroscience*, 34(34), 11233-11243.
- Diekelmann, S., & Born, J. (2010). The memory function of sleep. *Nature Reviews Neuroscience*, 11(2), 114-126.
- Doehrmann, O., & Naumer, M. J. (2008). Semantics and the multisensory brain: how meaning modulates processes of audio-visual integration. *Brain Research*, 1242, 136-150.
- Driver, J., & Noesselt, T. (2008). Multisensory interplay reveals crossmodal influences on 'sensory-specific' brain regions, neural responses, and judgments. *Neuron*, 57(1), 11-23.
- Dudai, Y. (2012). The restless engram: consolidations never end. *Annual review of neuroscience*, 35, 227-247.
- Easton, A., & Gaffan, D. (2001). Crossed unilateral lesions of the medial forebrain bundle and either inferior temporal or frontal cortex impair

- object-reward association learning in Rhesus monkeys. *Neuropsychologia*, 39(1), 71-82.
- Eggermont, J. J., & Ponton, C. W. (2002). The neurophysiology of auditory perception: from single units to evoked potentials. *Audiology & Neuro-Otology*, 7(2), 71-99.
- Eklund, K., Torppa, M., Aro, M., Leppänen, P. H. T., & Lyytinen, H. (2015). Literacy skill development of children with familial risk for dyslexia through grades 2, 3, and 8. *Journal of Educational Psychology*, 107(1), 126-140.
- Elbro, C., Daugaard, H. T., & Gellert, A. S. (2012). Dyslexia in a second language?-a dynamic test of reading acquisition may provide a fair answer. *Annals of Dyslexia*, 62(3), 172-185.
- Folia, V., Uddén, J., De Vries, M., Forkstam, C., & Petersson, K. M. (2010). Artificial Language Learning in Adults and Children. *Language Learning*, 60, 188-220.
- Fort, A. (2002). Dynamics of Cortico-subcortical Cross-modal Operations Involved in Audio-visual Object Detection in Humans. *Cerebral Cortex*, 12(10), 1031-1039.
- Foxe, J. J., Morocz, I. A., Murray, M. M., Higgins, B. A., Javitt, D. C., & Schroeder, C. E. (2000). Multisensory auditory-somatosensory interactions in early cortical processing revealed by high-density electrical mapping. *Cognitive Brain Research*, 10(1-2), 77-83.
- Froyen, D., van Atteveldt, N., Bonte, M., & Blomert, L. (2008). Cross-modal enhancement of the MMN to speech-sounds indicates early and automatic integration of letters and speech-sounds. *Neuroscience Letters*, 430(1), 23-28.
- Froyen, D., Willems, G., & Blomert, L. (2011). Evidence for a specific cross-modal association deficit in dyslexia: an electrophysiological study of letter - speech sound processing. *Developmental science*, 14(4), 635-648.
- Froyen, D. J. W., Bonte, M. L., van Atteveldt, N., & Blomert, L. (2009). The long road to automation: neurocognitive development of letter-speech sound processing. *Journal of Cognitive Neuroscience*, 21(3), 567-580.
- Ghazanfar, A. A., & Schroeder, C. E. (2006). Is neocortex essentially multisensory? *Trends in Cognitive Sciences*, 10(6), 278-285.
- Gibson, J. R., & Maunsell, J. H. (1997). Sensory modality specificity of neural activity related to memory in visual cortex. *Journal of Neurophysiology*, 78(3), 1263-1275.
- 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, 267.
- Groppe, D. M., Urbach, T. P., & Kutas, M. (2011). Mass univariate analysis of event - related brain potentials/fields I: A critical tutorial review. *Psychophysiology*, 48(12), 1711-1725.

- Grunewald, A., Linden, J. F., & Andersen, R. A. (1999). Responses to auditory stimuli in macaque lateral intraparietal area I. Effects of training. *Journal of neurophysiology*, 82(1), 330-342.
- Hämäläinen, J. A., Lohvansuu, K., Ervast, L., & Leppänen, P. H. (2015). Event-related potentials to tones show differences between children with multiple risk factors for dyslexia and control children before the onset of formal reading instruction. *International Journal of Psychophysiology*, 95(2), 101-112.
- Hämäläinen, J. A., Parviainen, T., Hsu, Y.-F., & Salmelin, R. (2019). Dynamics of brain activation during learning of syllable-symbol paired associations. *Neuropsychologia*, 129, 93-103.
- Hämäläinen, M. S., & Ilmoniemi, R. J. (1994). Interpreting magnetic fields of the brain: minimum norm estimates. *Medical & Biological Engineering & Computing*, 32(1), 35-42.
- Hashimoto, R., & Sakai, K. L. (2004). Learning letters in adulthood: direct visualization of cortical plasticity for forming a new link between orthography and phonology. *Neuron*, 42(2), 311-322.
- Häyrynen, T., Serenius-Sirve, S., & Korkman, M. (1999). Lukilasse. Lukemisen, Kirjoittamisen Ja Laskemisen Seulontatestistö Peruskoulun Ala-Asteen Luokille, 1-6.
- Hebscher, M., Wing, E., Ryan, J., & Gilboa, A. (2019). Rapid Cortical Plasticity Supports Long-Term Memory Formation. *Trends in Cognitive Sciences*, 23(12), 989-1002.
- Hein, G., Doehrmann, O., Muller, N. G., Kaiser, J., Muckli, L., & Naumer, M. J. (2007). Object Familiarity and Semantic Congruency Modulate Responses in Cortical Audiovisual Integration Areas. *Journal of Neuroscience*, 27(30), 7881-7887.
- Herdman, A. T., Fujioka, T., Chau, W., Ross, B., Pantev, C., & Picton, T. W. (2006). Cortical oscillations related to processing congruent and incongruent grapheme-phoneme pairs. *Neuroscience Letters*, 399(1-2), 61-66.
- Herdman, A. T., & Takai, O. (2013). Paying attention to orthography: a visual evoked potential study. *Frontiers in Human Neuroscience*, 7, 199.
- Hickok, G. (2012). The cortical organization of speech processing: feedback control and predictive coding the context of a dual-stream model. *Journal of Communication Disorders*, 45(6), 393-402.
- Hickok, G., & Poeppel, D. (2000). Towards a functional neuroanatomy of speech perception. *Trends in Cognitive Sciences*, 4(4), 131-138.
- Hocking, J., & Price, C. J. (2009). Dissociating verbal and nonverbal audiovisual object processing. *Brain and Language*, 108(2), 89-96.
- Holloway, I. D., van Atteveldt, N., Blomert, L., & Ansari, D. (2015). Orthographic dependency in the neural correlates of reading: evidence from audiovisual integration in English readers. *Cerebral Cortex*, 25(6), 1544-1553.

- Hyvärinen, A. (1999). Fast and robust fixed-point algorithms for independent component analysis. *IEEE Transactions on Neural Networks*, 10(3), 626–634.
- Jarrard, L. E. (1993). On the role of the hippocampus in learning and memory in the rat. *Behavioral and Neural Biology*, 60(1), 9–26.
- Jones, J. A., & Callan, D. E. (2003). Brain activity during audiovisual speech perception: an fMRI study of the McGurk effect. *Neuroreport*, 14(8), 1129–1133.
- Jost, L. B., Eberhard-Moscicka, A. K., Frisch, C., Dellwo, V., & Maurer, U. (2014). Integration of spoken and written words in beginning readers: a topographic ERP study. *Brain Topography*, 27(6), 786–800.
- Kamke, M. R., Vieth, H. E., Cottrell, D., & Mattingley, J. B. (2012). Parietal disruption alters audiovisual binding in the sound-induced flash illusion. *NeuroImage*, 62(3), 1334–1341.
- Karipidis, I. I., Pleisch, G., Brandeis, D., Roth, A., Röthlisberger, M., Schneebeli, M., ... & Brem, S. (2018). Simulating reading acquisition: the link between reading outcome and multimodal brain signatures of letter–speech sound learning in prereaders. *Scientific reports*, 8(1), 1–13.
- Karipidis, I. I., Pleisch, G., Röthlisberger, M., Hofstetter, C., Dornbierer, D., Stämpfli, P., & Brem, S. (2017). Neural initialization of audiovisual integration in prereaders at varying risk for developmental dyslexia. *Human Brain Mapping*, 38(2), 1038–1055.
- Katzir, T., Misra, M., & Poldrack, R. A. (2005). Imaging phonology without print: assessing the neural correlates of phonemic awareness using fMRI. *NeuroImage*, 27(1), 106–115.
- Kayser, C., Petkov, C. I., Augath, M., & Logothetis, N. K. (2007). Functional imaging reveals visual modulation of specific fields in auditory cortex. *Journal of Neuroscience*, 27(8), 1824–1835.
- Kayser, C., Petkov, C. I., & Logothetis, N. K. (2008). Visual Modulation of Neurons in Auditory Cortex. *Cerebral Cortex*, 18(7), 1560–1574.
- Kirby, J. R., Georgiou, G. K., Martinussen, R., Parrila, R., Bowers, P., & Landerl, K. (2010). Naming Speed and Reading: From Prediction to Instruction. *Reading Research Quarterly*, 45(3), 341–362.
- Klinzing, J. G., Niethard, N., & Born, J. (2019). Mechanisms of systems memory consolidation during sleep. *Nature neuroscience*, 22(10), 1598–1610.
- Korkman, M., Kirk, U., & Kemp, S. (1998). A developmental NEUROPSYCHOLOGICAL assessment (NEPSY). San Antonio, TX: Psychological Corporation, 64, 156–157.
- Korkman, M., Kirk, U., & Kemp, S. (2007). NEPSY – Second Edition (NEPSY-II). San Antonio, TX: Harcourt Assessment.
- Kuo, W. J., Yeh, T. C., Duann, J. R., Wu, Y. T., Ho, L. T., Hung, D., ... & Hsieh, J. C. (2001). A left-lateralized network for reading Chinese words: a 3 T fMRI study. *Neuroreport*, 12(18), 3997–4001.
- Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences: brain potentials reflect semantic incongruity. *Science*, 207(4427), 203–205.

- Kuuluvainen, S., Leminen, A., & Kujala, T. (2016). Auditory evoked potentials to speech and nonspeech stimuli are associated with verbal skills in preschoolers. *Developmental Cognitive Neuroscience*, 19, 223–232.
- Lakatos, P., Chen, C.-M., O’Connell, M. N., Mills, A., & Schroeder, C. E. (2007). Neuronal oscillations and multisensory interaction in primary auditory cortex. *Neuron*, 53(2), 279–292.
- Laurienti, P. J., Perrault, T. J., Stanford, T. R., Wallace, M. T., & Stein, B. E. (2005). On the use of superadditivity as a metric for characterizing multisensory integration in functional neuroimaging studies. *Experimental Brain Research*, 166(3-4), 289-297.
- Lerma-Usabiaga, G., Carreiras, M., & Paz-Alonso, P. M. (2018). Converging evidence for functional and structural segregation within the left ventral occipitotemporal cortex in reading. *Proceedings of the National Academy of Sciences*, 115(42), E9981-E9990.
- Lewis, J. W., Beauchamp, M. S., & DeYoe, E. A. (2000). A comparison of visual and auditory motion processing in human cerebral cortex. *Cerebral Cortex*, 10(9), 873–888.
- Liberman, A. M. (1992). The relation of speech to reading and writing. In *Advances in psychology* (Vol. 94, pp. 167-178). North-Holland.
- Lin, F.-H., Witzel, T., Ahlfors, S. P., Stufflebeam, S. M., Belliveau, J. W., & Hämäläinen, M. S. (2006). Assessing and improving the spatial accuracy in MEG source localization by depth-weighted minimum-norm estimates. *NeuroImage*, 31(1), 160–171.
- Lohvansuu, K., Hämäläinen, J. A., Ervast, L., Lyytinen, H., & Leppänen, P. H. T. (2018). Longitudinal interactions between brain and cognitive measures on reading development from 6 months to 14 years. *Neuropsychologia*, 108, 6–12.
- Lohvansuu, K., Hämäläinen, J. A., Tanskanen, A., Ervast, L., Heikkinen, E., Lyytinen, H., & Leppänen, P. H. (2014). Enhancement of brain event-related potentials to speech sounds is associated with compensated reading skills in dyslexic children with familial risk for dyslexia. *International journal of psychophysiology*, 94(3), 298-310.
- Lütkenhöner, B., Lammertmann, C., Simões, C., & Hari, R. (2002). Magnetoencephalographic correlates of audiotactile interaction. *NeuroImage*, 15(3), 509–522.
- Macaluso, E., & Driver, J. (2005). Multisensory spatial interactions: a window onto functional integration in the human brain. *Trends in Neurosciences*, 28(5), 264–271.
- Macaluso, E., Frith, C. D., & Driver, J. (2000). Modulation of human visual cortex by crossmodal spatial attention. *Science*, 289(5482), 1206-1208.
- Madec, S., Le Goff, K., Anton, J. L., Longcamp, M., Velay, J. L., Nazarian, B., ... & Rey, A. (2016). Brain correlates of phonological recoding of visual symbols. *NeuroImage*, 132, 359-372.
- Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods*, 164(1), 177–190.

- Martin, A., & Chao, L. L. (2001). Semantic memory and the brain: structure and processes. *Current Opinion in Neurobiology*, 11(2), 194–201.
- Maurer, U., Blau, V. C., Yoncheva, Y. N., & McCandliss, B. D. (2010). Development of visual expertise for reading: rapid emergence of visual familiarity for an artificial script. *Developmental Neuropsychology*, 35(4), 404–422.
- Maurer, U., Brem, S., Bucher, K., & Brandeis, D. (2005). Emerging neurophysiological specialization for letter strings. *Journal of Cognitive Neuroscience*, 17(10), 1532–1552.
- Mayes, A., Montaldi, D., & Migo, E. (2007). Associative memory and the medial temporal lobes. *Trends in Cognitive Sciences*, 11(3), 126–135.
- McCandliss, B. D., & Noble, K. G. (2003). The development of reading impairment: a cognitive neuroscience model. *Mental Retardation and Developmental Disabilities Research Reviews*, 9(3), 196–204.
- McClelland, J. L., McNaughton, B. L., & O'Reilly, R. C. (1995). Why there are complementary learning systems in the hippocampus and neocortex: insights from the successes and failures of connectionist models of learning and memory. *Psychological Review*, 102(3), 419–457.
- McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, 264(5588), 746–748.
- McNorgan, C., Randazzo-Wagner, M., & Booth, J. R. (2013). Cross-modal integration in the brain is related to phonological awareness only in typical readers, not in those with reading difficulty. *Frontiers in Human Neuroscience*, 7, 388.
- Mei, L., Xue, G., Lu, Z. L., He, Q., Wei, M., Zhang, M., ... & Chen, C. (2015). Native language experience shapes neural basis of addressed and assembled phonologies. *NeuroImage*, 114, 38–48.
- Mei, L., Xue, G., Lu, Z. L., He, Q., Zhang, M., Wei, M., ... & Dong, Q. (2014). Artificial language training reveals the neural substrates underlying addressed and assembled phonologies. *Plos one*, 9(3), e93548.
- Meredith, M. A. (2002). On the neuronal basis for multisensory convergence: a brief overview. *Brain Research. Cognitive Brain Research*, 14(1), 31–40.
- Meredith, M. A., & Stein, B. E. (1983). Interactions among converging sensory inputs in the superior colliculus. *Science*, 221(4608), 389–391.
- Meredith, M. A., & Stein, B. E. (1986). Visual, auditory, and somatosensory convergence on cells in superior colliculus results in multisensory integration. *Journal of Neurophysiology*, 56(3), 640–662.
- Mishkin, M., & Ungerleider, L. G. (1982). Contribution of striate inputs to the visuospatial functions of parieto-occipital cortex in monkeys. *Behavioural Brain Research*, 6(1), 57–77.
- Miyashita, Y., & Hayashi, T. (2000). Neural representation of visual objects: encoding and top-down activation. *Current Opinion in Neurobiology*, 10(2), 187–194.

- Moerel, M., De Martino, F., & Formisano, E. (2014). An anatomical and functional topography of human auditory cortical areas. *Frontiers in Neuroscience*, 8, 225.
- Molholm, S., Ritter, W., Murray, M. M., Javitt, D. C., Schroeder, C. E., & Foxe, J. J. (2002). Multisensory auditory-visual interactions during early sensory processing in humans: a high-density electrical mapping study. *Cognitive Brain Research*, 14(1), 115-128.
- Molholm, S., Sehatpour, P., Mehta, A. D., Shpaner, M., Gomez-Ramirez, M., Ortigue, S., ... & Foxe, J. J. (2006). Audio-visual multisensory integration in superior parietal lobule revealed by human intracranial recordings. *Journal of neurophysiology*, 96(2), 721-729.
- Moll, K., Ramus, F., Bartling, J., Bruder, J., Kunze, S., Neuhoff, N., ... & Tóth, D. (2014). Cognitive mechanisms underlying reading and spelling development in five European orthographies. *Learning and Instruction*, 29, 65-77.
- Murray, M. M., Molholm, S., Michel, C. M., Heslenfeld, D. J., Ritter, W., Javitt, D. C., ... & Foxe, J. J. (2005). Grabbing your ear: rapid auditory-somatosensory multisensory interactions in low-level sensory cortices are not constrained by stimulus alignment. *Cerebral cortex*, 15(7), 963-974.
- Murray, M. M., & Spierer, L. (2009). Auditory spatio-temporal brain dynamics and their consequences for multisensory interactions in humans. *Hearing Research*, 258(1-2), 121-133.
- Näätänen, R., Gaillard, A. W., & Mäntysalo, S. (1978). Early selective-attention effect on evoked potential reinterpreted. *Acta Psychologica*, 42(4), 313-329.
- Näätänen, R., Tervaniemi, M., Sussman, E., Paavilainen, P., & Winkler, I. (2001). 'Primitive intelligence' in the auditory cortex. *Trends in neurosciences*, 24(5), 283-288.
- Noesselt, T., Rieger, J. W., Schoenfeld, M. A., Kanowski, M., Hinrichs, H., Heinze, H. J., & Driver, J. (2007). Audiovisual temporal correspondence modulates human multisensory superior temporal sulcus plus primary sensory cortices. *Journal of Neuroscience*, 27(42), 11431-11441.
- Obleser, J., Lahiri, A., & Eulitz, C. (2003). Auditory-evoked magnetic field codes place of articulation in timing and topography around 100 milliseconds post syllable onset. *NeuroImage*, 20(3), 1839-1847.
- Ojanen, V., Möttönen, R., Pekkola, J., Jääskeläinen, I. P., Joensuu, R., Autti, T., & Sams, M. (2005). Processing of audiovisual speech in Broca's area. *NeuroImage*, 25(2), 333-338.
- Olson, I. R., Moore, K. S., Stark, M., & Chatterjee, A. (2006). Visual working memory is impaired when the medial temporal lobe is damaged. *Journal of Cognitive Neuroscience*, 18(7), 1087-1097.
- Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J.-M. (2011). FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Computational Intelligence and Neuroscience*, 2011, 156869.

- Parviainen, T., Helenius, P., Poskiparta, E., Niemi, P., & Salmelin, R. (2006). Cortical sequence of word perception in beginning readers. *Journal of Neuroscience*, 26(22), 6052-6061.
- Parviainen, T., Helenius, P., Poskiparta, E., Niemi, P., & Salmelin, R. (2011). Speech perception in the child brain: cortical timing and its relevance to literacy acquisition. *Human Brain Mapping*, 32(12), 2193-2206.
- Perrault Jr, T. J., Vaughan, J. W., Stein, B. E., & Wallace, M. T. (2005). Superior colliculus neurons use distinct operational modes in the integration of multisensory stimuli. *Journal of neurophysiology*, 93(5), 2575-2586.
- Pleisch, G., Karipidis, I. I., Brauchli, C., Röthlisberger, M., Hofstetter, C., Stämpfli, P., ... & Brem, S. (2019). Emerging neural specialization of the ventral occipitotemporal cortex to characters through phonological association learning in preschool children. *NeuroImage*, 189, 813-831.
- Plewko, J., Chyl, K., Bola, Ł., Łuniewska, M., Dębska, A., Banaszekiewicz, A., ... & Jednoróg, K. (2018). Letter and speech sound association in emerging readers with familial risk of dyslexia. *Frontiers in human neuroscience*, 12, 393.
- Price, C. J. (2000). The anatomy of language: contributions from functional neuroimaging. *Journal of anatomy*, 197(3), 335-359.
- Price, C. J., & Devlin, J. T. (2011). The interactive account of ventral occipitotemporal contributions to reading. *Trends in Cognitive Sciences*, 15(6), 246-253.
- Pugh, K. R., Landi, N., Preston, J. L., Mencl, W. E., Austin, A. C., Sibley, D., ... & Molfese, P. (2013). The relationship between phonological and auditory processing and brain organization in beginning readers. *Brain and language*, 125(2), 173-183.
- Pugh, K. R., Mencl, W. E., Jenner, A. R., Katz, L., Frost, S. J., Lee, J. R., ... & Shaywitz, B. A. (2000). Functional neuroimaging studies of reading and reading disability (developmental dyslexia). *Mental retardation and developmental disabilities research reviews*, 6(3), 207-213.
- Pugh, K. R., Mencl, W. E., Jenner, A. R., Katz, L., Frost, S. J., Lee, J. R., ... & Shaywitz, B. A. (2001). Neurobiological studies of reading and reading disability. *Journal of communication disorders*, 34(6), 479-492.
- Pugh, K. R., Mencl, W. E., Shaywitz, B. A., Shaywitz, S. E., Fulbright, R. K., Constable, R. T., ... & Liberman, A. M. (2000). The angular gyrus in developmental dyslexia: task-specific differences in functional connectivity within posterior cortex. *Psychological science*, 11(1), 51-56.
- Quak, M., London, R. E., & Talsma, D. (2015). A multisensory perspective of working memory. *Frontiers in Human Neuroscience*, 9, 197.
- Quinn, C., Taylor, J. S. H., & Davis, M. H. (2017). Learning and retrieving holistic and componential visual-verbal associations in reading and object naming. *Neuropsychologia*, 98, 68-84.
- Raij, T., Ahveninen, J., Lin, F. H., Witzel, T., Jääskeläinen, I. P., Letham, B., ... & Hämäläinen, M. (2010). Onset timing of cross - sensory activations and

- multisensory interactions in auditory and visual sensory cortices. *European Journal of Neuroscience*, 31(10), 1772-1782.
- Raij, T., Uutela, K., & Hari, R. (2000). Audiovisual integration of letters in the human brain. *Neuron*, 28(2), 617-625.
- Raschle, N. M., Chang, M., & Gaab, N. (2011). Structural brain alterations associated with dyslexia predate reading onset. *NeuroImage*, 57(3), 742-749.
- Raschle, N. M., Zuk, J., & Gaab, N. (2012). Functional characteristics of developmental dyslexia in left-hemispheric posterior brain regions predate reading onset. *Proceedings of the National Academy of Sciences*, 109(6), 2156-2161.
- Rüsseler, J., Ye, Z., Gerth, I., Szycik, G. R., & Münte, T. F. (2018). Audio-visual speech perception in adult readers with dyslexia: an fMRI study. *Brain imaging and behavior*, 12(2), 357-368.
- Ruzich, E., Crespo - García, M., Dalal, S. S., & Schneiderman, J. F. (2019). Characterizing hippocampal dynamics with MEG: A systematic review and evidence - based guidelines. *Human brain mapping*, 40(4), 1353-1375.
- Sakai, K., & Miyashita, Y. (1991). Neural organization for the long-term memory of paired associates. *Nature*, 354(6349), 152-155.
- Sandak, R., Mencl, W. E., Frost, S. J., & Pugh, K. R. (2004). The neurobiological basis of skilled and impaired reading: Recent findings and new directions. *Scientific Studies of reading*, 8(3), 273-292.
- Schlaggar, B. L., & McCandliss, B. D. (2007). Development of neural systems for reading. *Annual Review of Neuroscience*, 30, 475-503.
- Schroeder, C. E., & Foxe, J. J. (2002). The timing and laminar profile of converging inputs to multisensory areas of the macaque neocortex. *Brain Research. Cognitive Brain Research*, 14(1), 187-198.
- Schröger, E., & Widmann, A. (1998). Speeded responses to audiovisual signal changes result from bimodal integration. *Psychophysiology*, 35(6), 755-759.
- Scolari, M., Seidl-Rathkopf, K. N., & Kastner, S. (2015). Functions of the human frontoparietal attention network: Evidence from neuroimaging. *Current Opinion in Behavioral Sciences*, 1, 32-39.
- Sekiyama, K., Kanno, I., Miura, S., & Sugita, Y. (2003). Auditory-visual speech perception examined by fMRI and PET. *Neuroscience Research*, 47(3), 277-287.
- Shah-Basak, P. P., Urbain, C., Wong, S., da Costa, L., Pang, E. W., Dunkley, B. T., & Taylor, M. J. (2018). Concussion Alters the Functional Brain Processes of Visual Attention and Working Memory. *Journal of Neurotrauma*, 35(2), 267-277.
- Shulman, G. L., Pope, D. L., Astafiev, S. V., McAvoy, M. P., Snyder, A. Z., & Corbetta, M. (2010). Right hemisphere dominance during spatial selective attention and target detection occurs outside the dorsal frontoparietal network. *Journal of Neuroscience*, 30(10), 3640-3651.

- Simos, P. G., Rezaie, R., Fletcher, J. M., & Papanicolaou, A. C. (2013). Time-constrained functional connectivity analysis of cortical networks underlying phonological decoding in typically developing school-aged children: a magnetoencephalography study. *Brain and Language*, 125(2), 156-164.
- Song, Y., Hu, S., Li, X., Li, W., & Liu, J. (2010). The role of top-down task context in learning to perceive objects. *Journal of Neuroscience*, 30(29), 9869-9876.
- Specht, K., Hugdahl, K., Ofte, S., Nygård, M., Bjørnerud, A., Plante, E., & Helland, T. (2009). Brain activation on pre - reading tasks reveals at - risk status for dyslexia in 6 - year - old children. *Scandinavian Journal of Psychology*, 50(1), 79-91.
- Sperdin, H. F., Cappe, C., Foxe, J. J., & Murray, M. M. (2009). Early, low-level auditory-somatosensory multisensory interactions impact reaction time speed. *Frontiers in Integrative Neuroscience*, 3, 2.
- Stein, B. E. (1978). Development and organization of multimodal representation in cat superior colliculus. *Federation Proceedings*, 37(9), 2240-2245.
- Stein, B. E., & Stanford, T. R. (2008). Multisensory integration: current issues from the perspective of the single neuron. *Nature Reviews. Neuroscience*, 9(4), 255-266.
- Stekelenburg, J. J., Keetels, M., & Vroomen, J. (2018). Multisensory integration of speech sounds with letters vs. visual speech: only visual speech induces the mismatch negativity. *The European Journal of Neuroscience*, 47(9), 1135-1145.
- Stevens, C., McIlraith, A., Rusk, N., Niermeyer, M., & Waller, H. (2013). Relative laterality of the N170 to single letter stimuli is predicted by a concurrent neural index of implicit processing of letternames. *Neuropsychologia*, 51(4), 667-674.
- Stevenson, R. A., & James, T. W. (2009). Audiovisual integration in human superior temporal sulcus: Inverse effectiveness and the neural processing of speech and object recognition. *NeuroImage*, 44(3), 1210-1223.
- Sumby, W. H., & Pollack, I. (1954). Visual Contribution to Speech Intelligibility in Noise. *The Journal of the Acoustical Society of America*, 26(2), 212-215.
- Szymanski, M. D., Rowley, H. A., & Roberts, T. P. (1999). A hemispherically asymmetrical MEG response to vowels. *Neuroreport*, 10(12), 2481-2486.
- Tan, L. H., Liu, H. L., Perfetti, C. A., Spinks, J. A., Fox, P. T., & Gao, J. H. (2001). The neural system underlying Chinese logograph reading. *NeuroImage*, 13(5), 836-846.
- Tan, L. H., Spinks, J. A., Gao, J. H., Liu, H. L., Perfetti, C. A., Xiong, J., ... & Fox, P. T. (2000). Brain activation in the processing of Chinese characters and words: a functional MRI study. *Human brain mapping*, 10(1), 16-27.
- Tanabe, H. C., Honda, M., & Sadato, N. (2005). Functionally segregated neural substrates for arbitrary audiovisual paired-association learning. *Journal of Neuroscience*, 25(27), 6409-6418.

- Taulu, S., & Simola, J. (2006). Spatiotemporal signal space separation method for rejecting nearby interference in MEG measurements. *Physics in Medicine and Biology*, 51(7), 1759–1768.
- Taylor, J. S. H., Davis, M. H., & Rastle, K. (2017). Comparing and validating methods of reading instruction using behavioural and neural findings in an artificial orthography. *Journal of Experimental Psychology: General*, 146(6), 826.
- Taylor, J. S. H., Davis, M. H., & Rastle, K. (2019). Mapping visual symbols onto spoken language along the ventral visual stream. *Proceedings of the National Academy of Sciences*, 116(36), 17723–17728.
- Taylor, J. S. H., Rastle, K., & Davis, M. H. (2014). Distinct neural specializations for learning to read words and name objects. *Journal of Cognitive Neuroscience*, 26(9), 2128–2154.
- Taylor, M. J., Donner, E. J., & Pang, E. W. (2012). fMRI and MEG in the study of typical and atypical cognitive development. *Neurophysiologie Clinique/Clinical Neurophysiology*, 42(1-2), 19–25.
- Teder-Sälejärvi, W. A., McDonald, J. J., Di Russo, F., & Hillyard, S. A. (2002). An analysis of audio-visual crossmodal integration by means of event-related potential (ERP) recordings. *Brain Research. Cognitive Brain Research*, 14(1), 106–114.
- Torgesen, J. K., Rashotte, C. A., & Wagner, R. K. (1999). TOWRE: Test of word reading efficiency. Pro-Ed Austin, TX.
- van Atteveldt, N., & Ansari, D. (2014). How symbols transform brain function: A review in memory of Leo Blomert. *Trends in Neuroscience and Education*, 3(2), 44–49.
- van Atteveldt, N., Formisano, E., Goebel, R., & Blomert, L. (2004). Integration of letters and speech sounds in the human brain. *Neuron*, 43(2), 271–282.
- van Atteveldt, N., Roebroek, A., & Goebel, R. (2009). Interaction of speech and script in human auditory cortex: insights from neuro-imaging and effective connectivity. *Hearing research*, 258(1-2), 152–164.
- van Atteveldt, N. M., Formisano, E., Blomert, L., & Goebel, R. (2007). The effect of temporal asynchrony on the multisensory integration of letters and speech sounds. *Cerebral Cortex*, 17(4), 962–974.
- van Atteveldt, N. M., Formisano, E., Goebel, R., & Blomert, L. (2007). Top-down task effects overrule automatic multisensory responses to letter-sound pairs in auditory association cortex. *NeuroImage*, 36(4), 1345–1360.
- Vinckier, F., Dehaene, S., Jobert, A., Dubus, J. P., Sigman, M., & Cohen, L. (2007). Hierarchical coding of letter strings in the ventral stream: dissecting the inner organization of the visual word-form system. *Neuron*, 55(1), 143–156.
- Wagner, A. D., Paré-Blagoev, E. J., Clark, J., & Poldrack, R. A. (2001). Recovering meaning: left prefrontal cortex guides controlled semantic retrieval. *Neuron*, 31(2), 329–338.

- Wang, S., Zhu, Z., Zhang, J. X., Wang, Z., Xiao, Z., Xiang, H., & Chen, H.-C. (2008). Broca's area plays a role in syntactic processing during Chinese reading comprehension. *Neuropsychologia*, 46(5), 1371-1378.
- Wechsler, D. (1991). *WISC-III: Wechsler intelligence scale for children: Manual*. Psychological Corporation.
- Wechsler, D. (2003). *Wechsler Intelligence Scale for Children-Fourth Edition*. San Antonio, TX: The Psychological Corporation.
- Wechsler, D. (2008). *Wechsler adult intelligence scale--Fourth Edition (WAIS--IV)*. San Antonio, TX: NCS Pearson, 22, 498.
- Wilson, S. M., Bautista, A., & McCarron, A. (2018). Convergence of spoken and written language processing in the superior temporal sulcus. *Neuroimage*, 171, 62-74.
- Wu, C.-Y., Ho, M.-H. R., & Chen, S.-H. A. (2012). A meta-analysis of fMRI studies on Chinese orthographic, phonological, and semantic processing. *NeuroImage*, 63(1), 381-391.
- Wunderlich, J. L., Cone-Wesson, B. K., & Shepherd, R. (2006). Maturation of the cortical auditory evoked potential in infants and young children. *Hearing Research*, 212(1-2), 185-202.
- Xu, W., Kolozsvári, O. B., Monto, S. P., & Hämäläinen, J. A. (2018). Brain responses to letters and speech sounds and their correlations with cognitive skills related to reading in children. *Frontiers in human neuroscience*, 12, 304.
- Xu, W., Kolozsvári, O. B., Oostenveld, R., Leppänen, P. H. T., & Hämäläinen, J. A. (2019). Audiovisual Processing of Chinese Characters Elicits Suppression and Congruency Effects in MEG. *Frontiers in Human Neuroscience*, 13, 18.
- Yantis, S. (2008). The Neural Basis of Selective Attention: Cortical Sources and Targets of Attentional Modulation. *Current Directions in Psychological Science*, 17(2), 86-90.
- Yoncheva, Y. N., Blau, V. C., Maurer, U., & McCandliss, B. D. (2010). Attentional focus during learning impacts N170 ERP responses to an artificial script. *Developmental Neuropsychology*, 35(4), 423-445.
- Yoncheva, Y. N., Zevin, J. D., Maurer, U., & McCandliss, B. D. (2010). Auditory selective attention to speech modulates activity in the visual word form area. *Cerebral Cortex*, 20(3), 622-632.
- Yonelinas, A. P. (2013). The hippocampus supports high-resolution binding in the service of perception, working memory and long-term memory. *Behavioural Brain Research*, 254, 34-44.
- Žarić, G., Fraga González, G., Tijms, J., van der Molen, M. W., Blomert, L., & Bonte, M. (2014). Reduced neural integration of letters and speech sounds in dyslexic children scales with individual differences in reading fluency. *PloS One*, 9(10), e110337.



ORIGINAL PAPERS

I

AUDIOVISUAL PROCESSING OF CHINESE CHARACTERS ELICITS SUPPRESSION AND CONGRUENCY EFFECTS IN MEG

by

Weiyong Xu, Orsolya Beatrix Kolozsvári, Robert Oostenveld, Paavo Herman Tapio
Leppänen and Jarmo Arvid Hämäläinen, 2019

Frontiers in Human Neuroscience, 13, 18

<https://doi.org/10.3389/fnhum.2019.00018>

Reproduced with kind permission by Frontiers.



Audiovisual Processing of Chinese Characters Elicits Suppression and Congruency Effects in MEG

Weiyong Xu^{1,2*}, Orsolya Beatrix Kolozsvári^{1,2}, Robert Oostenveld^{3,4},
Paavo Herman Tapio Leppänen^{1,2} and Jarmo Arvid Hämäläinen^{1,2}

¹Department of Psychology, University of Jyväskylä, Jyväskylä, Finland, ²Jyväskylä Centre for Interdisciplinary Brain Research, Department of Psychology, University of Jyväskylä, Jyväskylä, Finland, ³Donders Institute for Brain, Cognition and Behaviour, Radboud University, Nijmegen, Netherlands, ⁴NatMEG, Department of Clinical Neuroscience, Karolinska Institutet, Stockholm, Sweden

OPEN ACCESS

Edited by:

Martin Schürmann,
University of Nottingham,
United Kingdom

Reviewed by:

Milene Bonte,
Maastricht University, Netherlands
Huan Luo,
Peking University, China

*Correspondence:

Weiyong Xu
weiyong.w.xu@jyu.fi

Received: 10 September 2018

Accepted: 16 January 2019

Published: 06 February 2019

Citation:

Xu W, Kolozsvári OB, Oostenveld R,
Leppänen PHT and Hämäläinen JA
(2019) Audiovisual Processing of
Chinese Characters Elicits
Suppression and Congruency
Effects in MEG.
Front. Hum. Neurosci. 13:18.
doi: 10.3389/fnhum.2019.00018

Learning to associate written letters/characters with speech sounds is crucial for reading acquisition. Most previous studies have focused on audiovisual integration in alphabetic languages. Less is known about logographic languages such as Chinese characters, which map onto mostly syllable-based morphemes in the spoken language. Here we investigated how long-term exposure to native language affects the underlying neural mechanisms of audiovisual integration in a logographic language using magnetoencephalography (MEG). MEG sensor and source data from 12 adult native Chinese speakers and a control group of 13 adult Finnish speakers were analyzed for audiovisual suppression (bimodal responses vs. sum of unimodal responses) and congruency (bimodal incongruent responses vs. bimodal congruent responses) effects. The suppressive integration effect was found in the left angular and supramarginal gyri (205–365 ms), left inferior frontal and left temporal cortices (575–800 ms) in the Chinese group. The Finnish group showed a distinct suppression effect only in the right parietal and occipital cortices at a relatively early time window (285–460 ms). The congruency effect was only observed in the Chinese group in left inferior frontal and superior temporal cortex in a late time window (about 500–800 ms) probably related to modulatory feedback from multi-sensory regions and semantic processing. The audiovisual integration in a logographic language showed a clear resemblance to that in alphabetic languages in the left superior temporal cortex, but with activation specific to the logographic stimuli observed in the left inferior frontal cortex. The current MEG study indicated that learning of logographic languages has a large impact on the audiovisual integration of written characters with some distinct features compared to previous results on alphabetic languages.

Keywords: audiovisual integration, magnetoencephalography, auditory cortex, language learning, reading, Chinese characters

INTRODUCTION

Learning to read involves the integration of multisensory information (primarily from the auditory and visual modalities) and combining it with meaning. Multisensory integration, defined as modulation of brain responses by signals from multiple modalities, has been shown to be a dynamic and context-dependent process (van Atteveldt et al., 2014; Murray et al., 2016a). In natural audiovisual speech perception, it has been shown that complementary or correlated visual speech information could affect the auditory processing of speech by activating the perisylvian auditory speech regions through the ventral and dorsal visual streams (Campanella and Belin, 2007; Campbell, 2008; Bernstein and Liebenthal, 2014). Multisensory integration is also found within the visual cortices during early and late post-stimulus stages and could directly impact perception (Murray et al., 2016b; Kayser et al., 2017). Unlike spoken language, the ability to read is not hard-wired in the human brain through the evolution since written language is a recent cultural invention which has only existed for a few thousand years (Liberman, 1992). Consequently, it takes years of repetition and practice to form the long-term memory representations of audiovisual language objects, which would enable fluent readers to successfully automatize the integration of language-related auditory and visual sensory information (Froyen et al., 2009). A growing body of neuroimaging research has examined the neurophysiological basis of the letter-speech sound integration mainly in transparent alphabetic languages (Raij et al., 2000; van Atteveldt et al., 2004, 2009; Froyen et al., 2009). Interestingly, research using less transparent language such as English has found both similarities in the audiovisual integration effect compared with transparent languages and orthography dependent differences in the processing of irregular mappings of letter-speech sound combinations (Holloway et al., 2015). Therefore, an intriguing open question is about the audiovisual integration in other kinds of languages, for example character-speech processing in logographic languages such as Chinese. Understanding of such character-speech integration in logographic languages may provide more insights into the universal and language-specific brain circuits underlying audiovisual integration in reading acquisition.

Audiovisual paradigms, which typically consisted of auditory only (A), visual only (V), audiovisual congruent (AVC) and audiovisual incongruent (AVI) stimuli, were widely used in investigating brain mechanisms of multisensory interactions between the auditory and visual modalities (Raij et al., 2000; Murray and Spierer, 2009; van Atteveldt et al., 2009). Within the structure of such experimental design, two main approaches could be derived and used as indications of audiovisual integration. The first approach is based on the additive model, which compares the audiovisual responses to the summations of the constituent unisensory responses [$AV - (A + V)$], and has been frequently used in electrophysiological studies on multisensory integration (Raij et al., 2000; Calvert and Thesen, 2004; Stein and Stanford, 2008; Sperdin et al., 2009). The additive model could be applied to almost any kind of

audiovisual experimental design with arbitrary auditory and visual combinations. This approach is suitable to detect both supra-additive [$AV > (A + V)$] and sub-additive (hereon referred to as the suppression effect) [$AV < (A + V)$] modulations of unimodal activities in the sensory-specific cortices as well as to observe new processes specifically activated by the bimodal nature of the stimulus under the assumption that there is no common activity (such as target processing) presented in the auditory, visual and audiovisual conditions (Besle et al., 2004). Animal electrophysiological studies have shown both super- and sub-additive multisensory interactions in the superior colliculus neurons and the superior temporal sulcus (STS; Meredith, 2002; Schroeder and Foxe, 2002; Laurienti et al., 2005; Perrault et al., 2005; Kayser et al., 2008; Stein and Stanford, 2008). Electroencephalography (EEG)/magnetoencephalography (MEG) studies on humans have typically shown suppressive multisensory effects (Schröger and Widmann, 1998; Foxe et al., 2000; Raij et al., 2000; Fort et al., 2002; Lütkenhöner et al., 2002; Molholm et al., 2002; Teder-Sälejärvi et al., 2002; Jost et al., 2014; Xu et al., 2018). Such suppression effects could occur as early as 50–60 ms after the stimulus onset and these functionally coupled responses are localized within the primary visual and auditory cortices as well as the posterior STS (Cappe et al., 2010). Other research in humans has also demonstrated that these audiovisual suppression effects can be observed at late time windows and for both familiar and unfamiliar audiovisual stimuli (Raij et al., 2000; Jost et al., 2014; Xu et al., 2018).

The second approach is to study the audiovisual congruency effect (Jones and Callan, 2003; Ojanen et al., 2005; Hein et al., 2007; Rüsseler et al., 2018), which compares different brain responses to congruent and incongruent audiovisual pairs. The rationale is that the congruency effect can only be established when the unisensory inputs have been integrated successfully (van Atteveldt et al., 2007a,b). One advantage of the congruency comparison is that it is not sensitive to any additional non-sensory activity and thus has a clear statistical criterion. Research using transparent languages such as Dutch and Finnish has found that the congruent alphabetic audiovisual stimuli elicit a stronger brain response in the superior temporal cortex than the incongruent pairs (Raij et al., 2000; van Atteveldt et al., 2004). Based on earlier studies (Raij et al., 2000; Besle et al., 2004; Cappe et al., 2010; Jost et al., 2014), the suppression effect reflects general audiovisual integration (including both early and late audiovisual interaction effects), whereas the congruency effect is more related to the specific interaction of learned or meaningful audiovisual associations (Hocking and Price, 2009).

Previous research has mainly focused on the brain mechanisms of audiovisual integration in alphabetic languages such as Dutch (van Atteveldt et al., 2004, 2009), English (Holloway et al., 2015) and Finnish (Raij et al., 2000). Several language-related and cross-modal brain regions have been shown to activate consistently during letter-speech sound integration in alphabetic languages. In particular, the superior temporal cortices have been reported in fMRI studies to have heteromodal properties (van Atteveldt et al., 2004, 2009; Blau et al., 2008). The left and right STS have also been implicated to be the main letter-speech sound integration regions in

an early MEG study using Finnish letters (Raij et al., 2000). In addition, feedback projections from these cross-modal regions were found to alter the brain activities in the primary auditory cortex (van Atteveldt et al., 2004). The heteromodal areas in the temporal cortices have shown differences in their tolerances to temporal synchrony between modalities: the visual and auditory inputs are integrated in the STS and superior temporal gyrus (STG) within a broad range of temporal cross-modal synchrony between the auditory and visual stimuli, while the effect of congruent and incongruent auditory-visual stimuli rapidly diminishes with decreasing temporal synchrony in planum temporale (PT) and Heschl's sulcus (HS) regions (van Atteveldt et al., 2007a). Top-down influences by different instructions and task demands also evidently affect the congruency effects (Andersen et al., 2004). For instance, different experimental designs (explicit/implicit and active/passive) have been shown to modulate the letter-speech sound congruency effect in fMRI (van Atteveldt et al., 2007b; Blau et al., 2008).

While fMRI provides accurate locations of the integration sites, its poor temporal resolution fails to capture the timing information. EEG and MEG have an excellent temporal resolution (in millisecond scale) and could provide additional timing information about audiovisual integration processes. An early MEG study using Finnish letters and speech sounds revealed that the auditory and visual sensory inputs showed maximal activities in multimodal sites about 225 ms after stimulus onset (Raij et al., 2000). It was followed by suppressive interaction at 280–345 ms in the right temporo-occipitoparietal junction and at 380–540 ms in the left STS and 450–535 ms in the right STS (Raij et al., 2000). Another MEG study using Hiragana graphemes and phonemes found that congruent audiovisual stimuli evoked larger 2–10 Hz oscillations in the left auditory cortex within the first 250 ms after stimulus onset and smaller 2–16 Hz oscillations in bilateral visual cortices between 250 and 500 ms compared with incongruent audiovisual stimuli (Herdman et al., 2006). Corroborating evidence comes from an EEG study that found audiovisual suppression effects in event-related potentials (ERPs) for familiar German (300–324 and 480–764 ms) and unfamiliar English words (324–384 and 416–756 ms), whereas audiovisual congruency effect could be found only for familiar German words (160–204, 544–576, 1032–1108 and 1164–1188 ms), and it was characterized by topographic differences probably due to lexical-semantic processes involved (Jost et al., 2014).

However, most studies on audiovisual integration have only investigated letter-speech sound mappings in alphabetic languages, and less is known about the logographic language such as Chinese. In Chinese, a single character encodes morphosyllabic information that represents a syllable with a distinct meaning (Perfetti and Tan, 1998; Shu, 2003; Tan et al., 2005; Ziegler and Goswami, 2005; McBride, 2015). For native Chinese speakers, long-term memory representations of characters through years of learning and reading repetition would enable them to retrieve the phonological and semantic information embedded in the characters directly. Semantic

processing has been shown to modulate the brain activity around 400 ms, with the N400 response being sensitive to semantic incongruency (Kutas and Federmeier, 2011; Du et al., 2014; Jost et al., 2014). The magnetic counterpart of the N400 (N400m) has also been shown to be sensitive to semantic content and to have sources in the bilateral temporal cortices in unimodal conditions (Service et al., 2007; Vartiainen et al., 2009). Thus, research using logographic language such as Chinese is important to advance our understanding of the general mechanisms of audiovisual integration including cross-modal semantic congruency processing in the human brain.

In the current study, we investigated the dynamics of cortical activation to logographic multisensory stimuli using MEG. We designed an active one-back audiovisual cover task during which audiovisual integration could be examined. We presented Chinese characters and speech sounds as stimuli in A, V, AVC, and AVI conditions to native speakers of Chinese and used native speakers of Finnish, who were naive to Chinese, as a control group to verify the effects of long-term exposure to these stimuli. The MEG data were analyzed at both sensor and source levels to examine the suppression and congruency effects. Highly automatized audiovisual associations could affect processes related to general cross-modal integration manifested as suppression effect and congruency detection. More specifically, we hypothesized that learned characters combined with the corresponding congruent/incongruent speech sounds would elicit an audiovisual congruency effect manifested as a combined modulatory feedback and semantic N400m response, but only in native Chinese speakers. In addition, we examined the effects of long-term exposure of logographic language on the audiovisual suppressive interaction (sum of auditory and visual only conditions compared to audiovisual condition) which should be affected less by semantic processing than the congruency effect. Instead, suppressive integration is likely to reflect more general cross-modal processes. We thus expected more similar suppression than congruency effect to alphabetic languages found in earlier studies.

MATERIALS AND METHODS

Participants

Chinese participants were adults and native speakers of Mandarin Chinese studying in Jyväskylä, Finland. Chinese participants had learned simplified Chinese characters through formal education. In addition, another group of native speakers of Finnish was recruited as a control group. Finnish participants had no prior exposure to the Chinese language. All participants included had normal hearing and normal or corrected-to-normal vision. They were screened for the following exclusion criteria: head injuries, ADHD, neurological diseases, medication affecting the central nervous system, language delays or any other language-related disorders. Ethical approval for this study was obtained from the Ethics Committee of the University of Jyväskylä. This study was carried out in accordance with the recommendations of the Ethics Committee of the University of Jyväskylä with written informed consent from

all subjects. All subjects gave written informed consent in accordance with the Declaration of Helsinki. The protocol was approved by the Ethics Committee of the University of Jyväskylä. After the MEG experiments, all of them received movie tickets as compensations for their time. In total, we measured MEG data from 19 native Chinese speakers and 18 Finnish speakers. Of those seven were excluded from the Chinese group and five were excluded in the Finnish group for following reasons: four subjects due to poor visual acuity even with magnetically neutral glasses for vision correction, two subjects due to excessive head movements or low head positions in the MEG helmet, one due to technical problems during the recording, and five subjects due to strong noise interference and poor signal quality. The final dataset thus included 12 Chinese participants and 13 Finnish participants (Table 1).

Stimuli and Task

The stimuli consisted of six Chinese characters (Simplified Chinese) and their corresponding flat tone speech sounds (1. 步: bu; 2. 都: du; 3. 谷: gu; 4. 酷: ku; 5. 普: pu; 6. 兔: tu). The characters were all common characters familiar to the native Chinese speakers and had the following meanings: 1. steps/walking; 2. both/capital; 3. grain; 4. cool; 5. common; and 6. rabbit. The mean duration of the auditory stimuli was 447.2 ms (SD: 32.7 ms). The duration of the visual stimuli was 1,000 ms. Four kinds of stimuli, A, V, AVC and AVI were presented in random order with 108 trials in each type of the stimuli. Each trial started with a fixation cross at the center of the screen for 1,000 ms. In the AVI and AVC conditions, there was a 36 ms delay between the visual and auditory stimuli. This time delay is within the optimal range of cross-modal integration (Schroeder and Foxe, 2002; Kayser et al., 2008). The visual stimuli were projected onto the center of the screen one meter away from the participants with a white color background. The fixation cross was 1.2 cm and the characters were 2.5 × 2.5 cm on the screen. The stimuli were presented with Presentation (Neurobehavioral Systems, Inc., Albany, CA, USA) software running in a Microsoft Windows computer.

Participants were instructed to do a dual-modality one-back task in order to keep their attention equally on both auditory and visual stimuli (Figure 1). Cover task trials as shown in Figure 1 occurred randomly (with 7.5% probability across all trials) to keep their attention on both auditory and visual modalities. The cover task trials tested the participant’s memory about the auditory and visual stimuli one trial before the last trial. The cover task trials consisted of one test trial for the auditory stimulus, one test trial for the visual stimulus followed by feedback. Four options were given on the screen in each

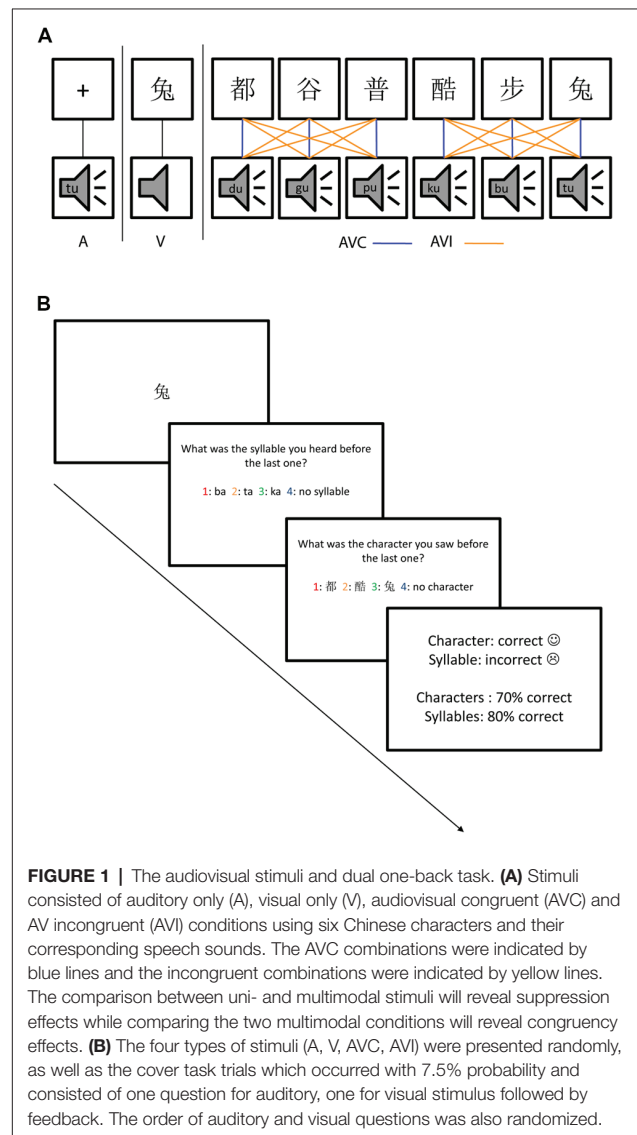


FIGURE 1 | The audiovisual stimuli and dual one-back task. **(A)** Stimuli consisted of auditory only (A), visual only (V), audiovisual congruent (AVC) and AV incongruent (AVI) conditions using six Chinese characters and their corresponding speech sounds. The AVC combinations were indicated by blue lines and the incongruent combinations were indicated by yellow lines. The comparison between uni- and multimodal stimuli will reveal suppression effects while comparing the two multimodal conditions will reveal congruency effects. **(B)** The four types of stimuli (A, V, AVC, AVI) were presented randomly, as well as the cover task trials which occurred with 7.5% probability and consisted of one question for auditory, one for visual stimulus followed by feedback. The order of auditory and visual questions was also randomized.

of the auditory or visual test trials and the participants needed to choose the correct one from the four options using the response pad. The order of the auditory and visual test trials was randomized.

MEG Recording

MEG data (102 magnetometer channels and 204 planar gradiometer channels; sampling rate: 1,000 Hz; band-pass filter: 0.1–330 Hz) were recorded using Elekta Neuromag® TRIUX™ system (Elekta AB, Stockholm, Sweden) in a magnetically shielded room. The head position with respect to the MEG sensors in the helmet was monitored continuously with five digitized head position indicator (HPI) coils attached to the scalp. Three HPI coils were placed on the forehead and one behind each ear. At the beginning of each MEG recording, the positions of HPI coils were determined by three anatomic landmarks (nasion, left and right preauricular points), which were digitized using the

TABLE 1 | Demographic information of the study participants.

Group	Chinese	Finnish
Number of participants	12	13
Age	24.36 ± 3.66	24.31 ± 2.06
Sex	9 Female, 3 Male	9 Female, 4 Male
Handedness	12 right	12 right, 1 ambidextrous

Polhemus Isotrak digital tracker system (Polhemus, Colchester, VT, USA). To allow the co-registration with the MRI template, an additional set of points (>100) randomly distributed over the scalp were also digitized. Electro-oculogram (EOG) signal was recorded with one ground electrode attached to the collarbone and another two diagonally placed electrodes (one slightly above the right eye and one slightly below the left eye).

Data Analysis

First, MEG data were processed with Maxfilter 3.0™ (Elekta AB, Stockholm, Sweden) to remove external magnetic disturbance and correct for head movements (Taulu et al., 2004; Taulu and Kajola, 2005; Taulu and Simola, 2006). Bad channels were marked manually and were excluded and later reconstructed in the Maxfilter. The temporal extension of the signal-space separation method (tSSS) was applied in buffers of 30 s (Taulu et al., 2004; Taulu and Kajola, 2005; Taulu and Simola, 2006). The head position was estimated using 200 ms time windows with 10 ms steps for head movement compensation. The mean head position across the whole MEG recording session was used for head position transformation.

MEG data were then analyzed using MNE Python (0.14; Gramfort et al., 2013). First, a low-pass filter of 40 Hz (firwin2 filter design, transition bandwidth 10 Hz) was applied to the continuous MEG recordings and data were segmented into epochs -200 to 1,000 ms relative to the stimulus onset. Data were then manually checked to excluded any trials that contaminated by head movement related artifacts or electronic jump artifacts. Then fast independent component analysis (ICA) was applied to remove any eye blink and cardiac artifacts related components. MEG epochs exceeding 1 pT/cm (for gradiometer channels) or 3pT (for magnetometer channels) peak-to-peak amplitudes were excluded from further analysis. Event-related fields (ERFs) were obtained by averaging trials in the four conditions (A, V, AVC, and AVI) separately. Sum of auditory and visual responses (A + V) was calculated by adding up the auditory and visual ERFs

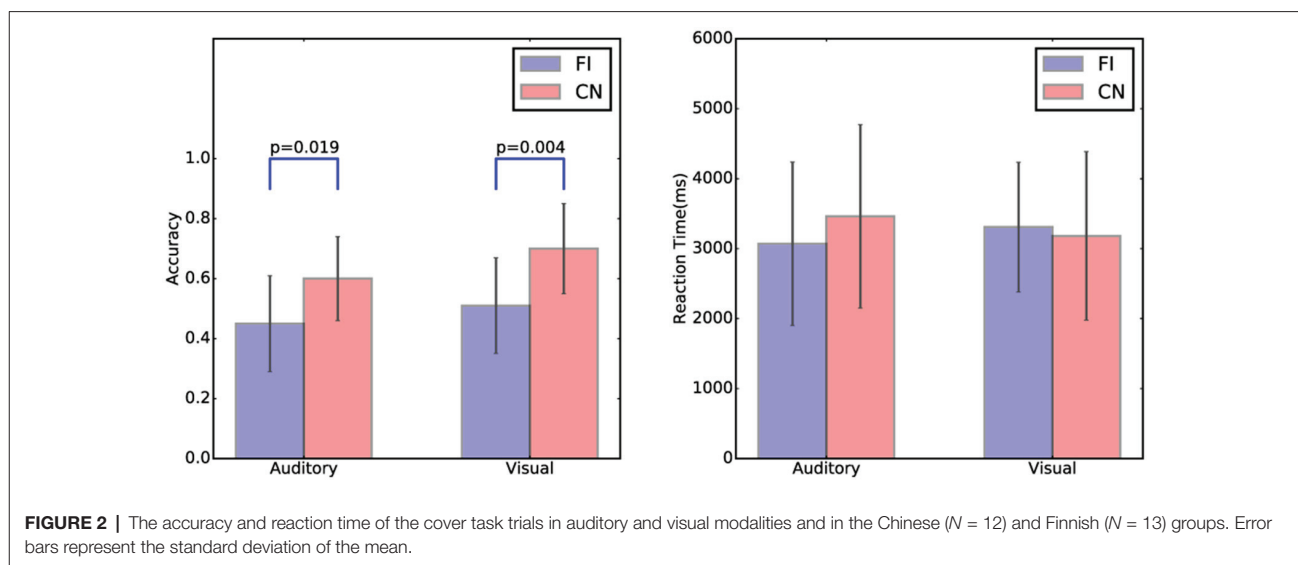
together (the numbers of trials in A and V conditions were equalized). To match the noise level of the A + V and AVC conditions and thus make these two conditions comparable, a subset of the AVC trials was created by randomly selecting about half the number of trials from the AVC condition as the number of trials in the A + V condition. For sensor level comparison, the gradiometer channel pairs in two orthogonal directions were combined using the vector sum method implemented in FieldTrip toolbox (Oostenveld et al., 2011). Gradiometers were chosen because they are less sensitive to noise sources originating far from the sensors than the magnetometers.

Individual magnetic resonance images (MRI) were not available from the participants and therefore the Freesurfer (RRID:SCR_001847) average brain (FSAverage) was used as a template for the source analysis. Three-parameter scaling was used to coregister FSAverage with the individual digitized head points. The average distance between the digitized head points and the scaled template scalp surface was 3.55 mm (0.25 mm SD).

Depth-weighted L2-minimum-norm estimate (wMNE; Hämäläinen and Ilmoniemi, 1994; Lin et al., 2006) was calculated for 10,242 free orientation current dipoles distribute on the cortical surface in each hemisphere. The noise covariance matrix was estimated from the raw MEG data from the 200 ms pre-stimulus baseline. The inverse solution was noise-normalized using dynamical statistical parametric maps (dSPM; Dale et al., 2000) for further statistical analysis. Since the FSAverage brain template that was used for all participants was only scaled to the subject-specific head size, the estimated brain activity could be directly compared in the statistical analyses without morphing to a common brain.

Statistical Analysis

Sensor level statistical analyses on the combined gradiometers were conducted using the nonparametric permutation test (Maris and Oostenveld, 2007) with spatial-temporal clustering based on *t*-test statistic implemented in the FieldTrip toolbox.



Source level analyses were similarly conducted using the nonparametric permutation *t*-test with spatio-temporal clustering in MNE Python. The time window was selected from 0 to 1,000 ms after the stimulus onset and the number of permutations was 1,024 in both sensor and source level statistical analyses. The cluster alpha was 0.05 for the incongruent–congruent comparison. For the AV – (A + V) comparison, a more conservative cluster alpha value of 0.005 was used due to the lower SNR since only half of the trials were averaged compared with the congruency contrast.

RESULTS

Behavioral Performance

The accuracy and reaction time of the cover task trials in both Chinese and Finnish groups are shown in **Figure 2**. There was a significant group difference in the accuracy for both auditory ($t_{(23)} = -2.52, p = 0.019$) and visual modality ($t_{(23)} = -3.20, p = 0.004$) between Chinese and Finnish groups, with the Chinese participants (Auditory: mean = 0.60, SD = 0.13 ; Visual: mean = 0.70, SD = 0.15) being more accurate than the

Finnish participants (Auditory: mean = 0.45, SD = 0.16; Visual: mean = 0.51, SD = 0.16) as expected. No significant differences in the reaction time were found between Chinese and Finnish groups.

Grand Average

Figure 3 gives an overview of the brain responses to the unimodal and bimodal audiovisual stimuli in sensor and source level in both Chinese and Finnish groups. **Figure 3A** shows the grand average evoked waveforms for the A, V, AVC and AVI conditions averaged over left and right temporal and occipital channels (vector sum of the paired orthogonal gradiometer channels) in the Chinese and Finnish groups. **Figure 3B** shows the magnetic field topography and dSPM source activations of the peak evoked responses at early (100–200 ms) and late (300–700 ms) time windows for each of the four conditions.

Suppression Effect [AVC – (A + V)]

Testing for suppression effects in the latency range from 0 to 1,000 ms post-stimulus in both groups separately at the sensor level, the cluster-based permutation test revealed significant

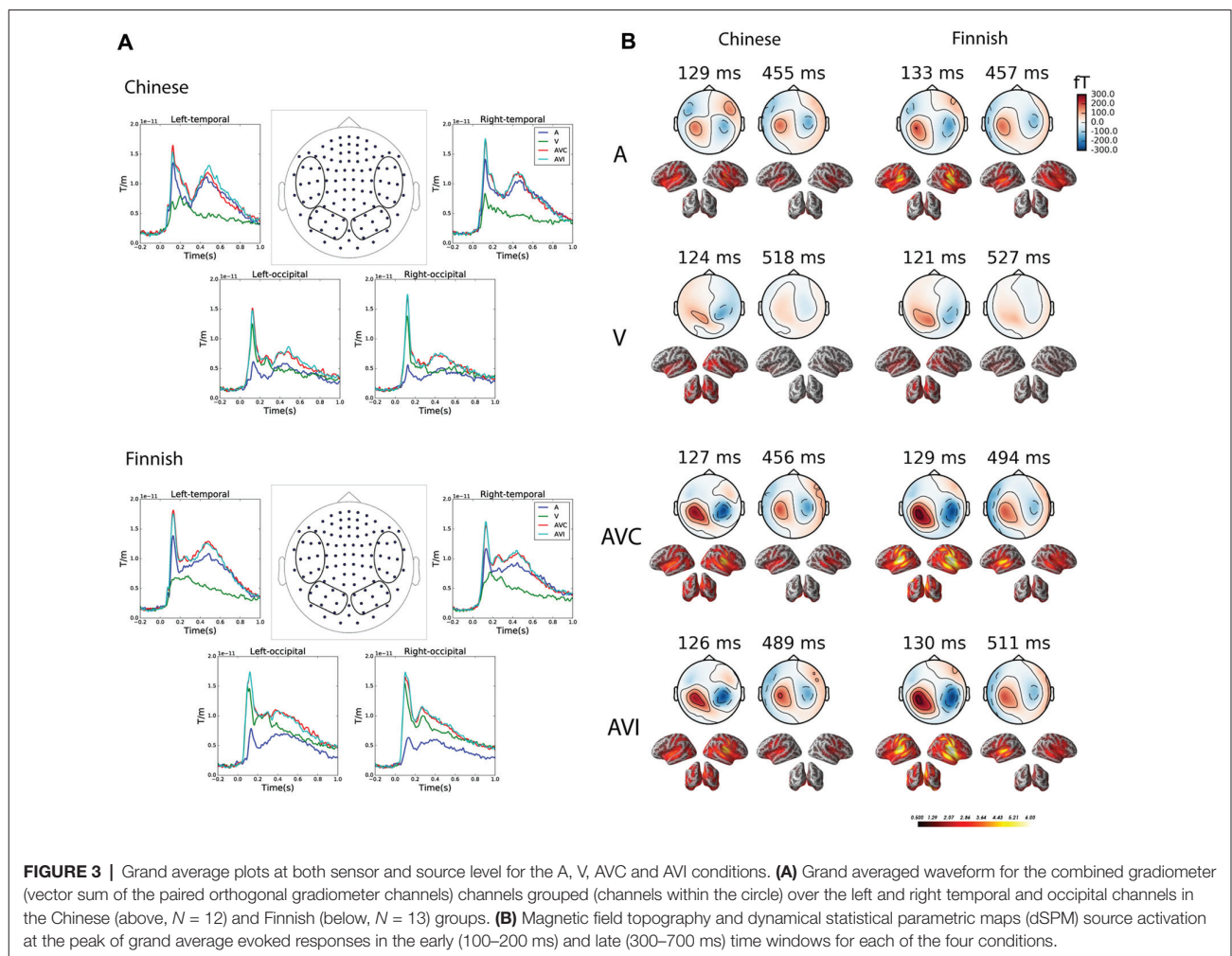


FIGURE 3 | Grand average plots at both sensor and source level for the A, V, AVC and AVI conditions. **(A)** Grand averaged waveform for the combined gradiometer (vector sum of the paired orthogonal gradiometer channels) channels grouped (channels within the circle) over the left and right temporal and occipital channels in the Chinese (above, $N = 12$) and Finnish (below, $N = 13$) groups. **(B)** Magnetic field topography and dynamical statistical parametric maps (dSPM) source activation at the peak of grand average evoked responses in the early (100–200 ms) and late (300–700 ms) time windows for each of the four conditions.

differences between the summed auditory and visual only conditions and the AVC conditions for both the Chinese group ($p = 0.002$) and the Finnish group ($p = 0.006$). In the Chinese group, the significant cluster was from 557 to 692 ms and mainly at the left temporal and frontal channels. In the Finnish group, the significant cluster was from 363 to 520 ms and mainly at the right parietal-occipital channels.

At the source level, the cluster-based permutation test also revealed significant differences between the summed auditory and visual only conditions and AVC conditions for both the Chinese group and the Finnish group. In the Chinese group, two significant clusters were found with different time windows and source locations. The first significant cluster was from 205 to 365 ms in the left angular and supramarginal gyri ($p = 0.01$). The second significant cluster was from 575 to 800 ms in the left temporal and frontal regions ($p = 0.001$). In the Finnish group, the significant cluster was from 285 to 460 ms and found in the right parietal-occipital regions ($p = 0.003$).

Results for both sensor and source level suppression integration comparisons and the spatiotemporal pattern of significant clusters are shown in **Figure 4**.

Congruency Effect (AVI – AVC)

At the sensor level, the congruency effect was tested in the latency range from 0 to 1,000 ms post-stimulus in both groups separately. The cluster-based permutation test revealed a significant difference between the congruent and incongruent conditions in the Chinese group only ($p = 0.01$; see **Figure 5**). The cluster occurred from 538 to 690 ms and the difference was most pronounced over the left frontal and temporal sensors in this latency range.

At the source level, the cluster-based permutation test also revealed a significant difference between the congruent and incongruent conditions in the Chinese group only ($p = 0.008$; see **Figure 5**). The cluster occurred from 490 to 890 ms and the difference was most pronounced over the left frontal and temporal regions.

DISCUSSION

In this study, we examined the effect of long-term exposure to Chinese characters on the audiovisual integration processes. We used a rigorous control of the participant’s attention on the two modalities by introducing a dual-modality one-back task. We observed that exposure to one’s native language and the long-term memory traces for spoken and written language indeed significantly changed the neural networks of audiovisual integration. This can be seen in the left-lateralized suppression effect, reflecting automatic audiovisual processes in the Chinese participants and a right-lateralized suppression effect in Finnish participants when processing novel Chinese audiovisual stimuli. Furthermore, long-term memory representations of Chinese characters are also manifested in how the brain reacts to the AVC and AVI stimuli in the left superior temporal and Broca’s area in native Chinese speakers.

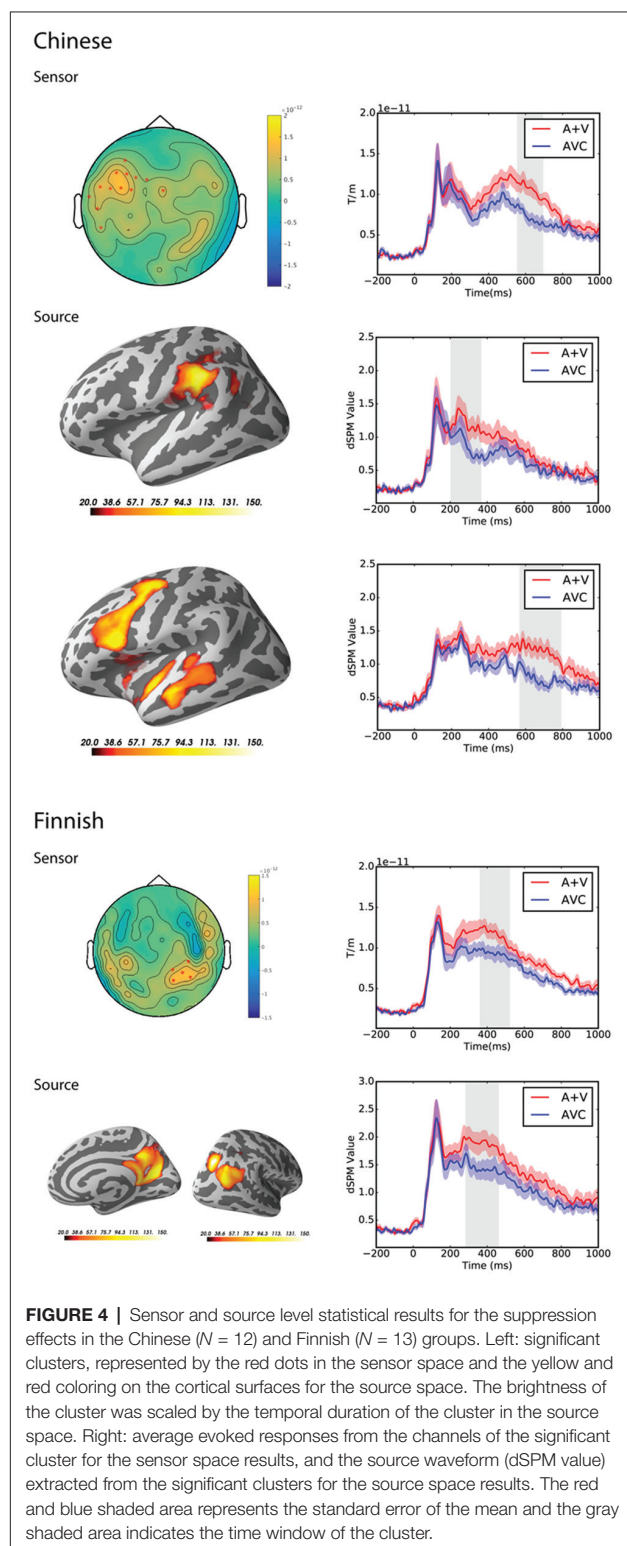


FIGURE 4 | Sensor and source level statistical results for the suppression effects in the Chinese ($N = 12$) and Finnish ($N = 13$) groups. Left: significant clusters, represented by the red dots in the sensor space and the yellow and red coloring on the cortical surfaces for the source space. The brightness of the cluster was scaled by the temporal duration of the cluster in the source space. Right: average evoked responses from the channels of the significant cluster for the sensor space results, and the source waveform (dSPM value) extracted from the significant clusters for the source space results. The red and blue shaded area represents the standard error of the mean and the gray shaded area indicates the time window of the cluster.

Suppression (AVC vs. A + V) effects were found in both the Chinese and Finnish groups but with different hemispheric lateralizations and at different time windows.

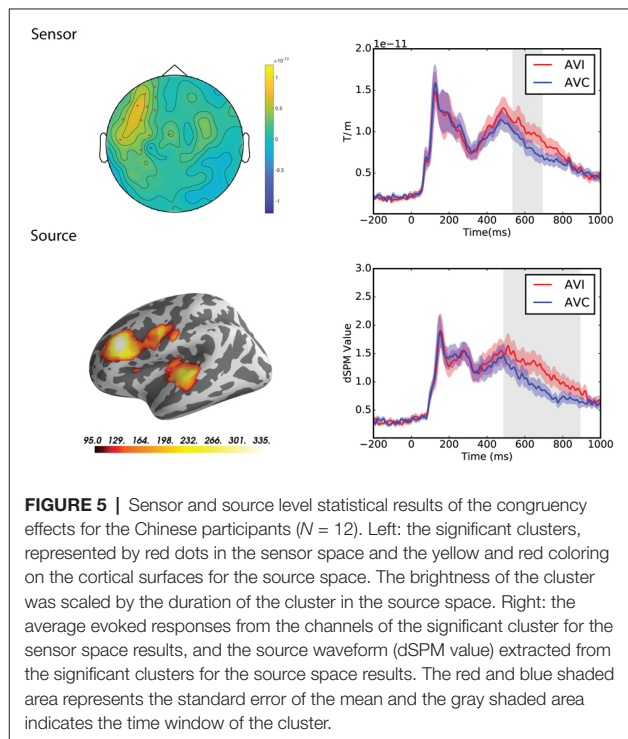


FIGURE 5 | Sensor and source level statistical results of the congruency effects for the Chinese participants ($N = 12$). Left: the significant clusters, represented by red dots in the sensor space and the yellow and red coloring on the cortical surfaces for the source space. The brightness of the cluster was scaled by the duration of the cluster in the source space. Right: the average evoked responses from the channels of the significant cluster for the sensor space results, and the source waveform (dSPM value) extracted from the significant clusters for the source space results. The red and blue shaded area represents the standard error of the mean and the gray shaded area indicates the time window of the cluster.

The suppression effect was most pronounced over the left angular/supramarginal, temporal and inferior frontal regions in the Chinese group, which suggested that the left hemisphere language network of native Chinese speakers was activated during processing of learned audiovisual association of Chinese characters and speech sounds. When children learn to read in an alphabetic language, visual letters and auditory phonemes are often presented simultaneously and neural pathways that map graphemes to phonemes are formed (Brem et al., 2010). Neural connectivity is strengthened through learning to read in school and later through reading practice, which would enable fluent readers to form long-term memory traces for written symbols and fast retrieval of audiovisual associations (Maurer et al., 2006). Consequently, the audiovisual suppression effect can be interpreted as the optimization of neural networks as a consequence of language learning in the Chinese participants (Raij et al., 2000). The early suppression effect was found in the left angular and supramarginal gyri about 200–350 ms after the stimulus onset. The left angular and supramarginal gyri have been considered as heteromodal areas related to linking orthographic representation from the occipital lobe to phonological coding represented in the STG (Price, 2000; Pugh et al., 2000; Schlaggar and McCandliss, 2007) and also possible feed forward connection to the inferior frontal gyrus (Simos et al., 2013). The left-lateralized suppression effect in the left superior and middle temporal areas in the later time window (about 550–800) matches those observed in earlier MEG, EEG and fMRI studies (Raij et al., 2000; Calvert et al., 2001; van Atteveldt et al., 2004) using alphabetic letters. The

effect in Broca's area is less often reported in audiovisual integration studies using alphabetic languages, possibly due to the fact that the Chinese language requires extra semantic processing in this area as shown by many other studies (Kuo et al., 2001; Tan et al., 2001; Wang et al., 2008). Taken together, it seems that the left superior temporal cortex is a common node in the neural network for audiovisual language processing for both alphabetic and logographic scripts. Importantly, the left inferior frontal cortex seems to be involved in the additional semantic-related audiovisual processing in logographic scripts (Wu et al., 2012; van Atteveldt and Ansari, 2014).

In the Finnish group, the suppression effect was most pronounced in the right inferior parietal and occipital area in a slightly earlier time window (285–460 ms), which partly matches with the findings by Raij et al. (2000) who also showed a predominant audiovisual interaction effect in the right temporo-occipito-parietal junction (280–345 ms). Clearly, for Finnish participants, Chinese characters did not have long-term memory representations and might be processed more like line drawings instead of meaningful characters. Parietal areas have been shown to be involved in multisensory processing in both human and monkey studies (Grunewald et al., 1999; Lewis et al., 2000; Bremmer et al., 2001). For example, lateral intraparietal area (LIP) which was originally considered as part of the unimodal visual cortex has been shown to respond to auditory stimuli in many studies (Bremmer et al., 2001; Ahveninen et al., 2006; Cohen, 2009). Other research has also indicated that the right parietal region plays a role in the perception of stimuli without any long-term representations arising simultaneously from the multiple sensory modalities (Kamke et al., 2012). The suppression effect in the right parietal and occipital regions at early time window indicates that the unfamiliar audiovisual information processing might rely more on the processing of the visual features (Calvert et al., 2001; Madec et al., 2016). It could be explained by the fact that Finnish participants who are naive to Chinese focused much more attention on the analysis of the spatial information of various strokes comprising the logographic character in order to be able to integrate the audiovisual stimuli. It should be noted that the suppression effect in the Finnish group was mostly related to this specific task and should not be interpreted as audiovisual integration related to logographic language. Therefore, the suppression effect found in the Finnish participants most likely represents a basic audiovisual processing (Calvert, 2001; Molholm et al., 2002; Cappe et al., 2010) of novel symbols in a working memory task.

Spatial-temporal clustering revealed an N400-like deflection elicited due to the congruency of the audiovisual stimuli in the Chinese, which was absent in the Finnish group. The congruency effect was mainly localized close to the left superior temporal cortex, Heschl's gyrus, inferior frontal cortex and also parts of the insula. The difference occurred in a late time window around 500–800 ms. This result is in line with previous studies using MEG and EEG (Raij et al., 2000; Jost et al., 2014). According to the functional neuroanatomical model for the integration of letters and

speech sounds proposed by van Atteveldt et al. (2009), the auditory and visual information is integrated in the heteromodal STS/STG and then feedback projected to the auditory cortex. In the current study, the congruency effect presented in the late time window therefore supports such feedback projection mechanism. Previous studies (Raij et al., 2000; van Atteveldt et al., 2004) using transparent alphabetic grapheme-phoneme associations found that the brain activations in the auditory association cortex were enhanced by congruent letters and suppressed by incongruent letters. This modulation is overruled during the explicit matching when both types of the audiovisual stimuli are equally relevant, independent of the congruency and temporal relation (van Atteveldt et al., 2007b). Whereas in opaque languages such as English, incongruent grapheme-phoneme associations have been shown to elicit a weaker and even reversed pattern of audiovisual integration to congruent pairs in the superior temporal areas (Holloway et al., 2015). Our results further indicated that audiovisual congruency is adaptive to different script types with a reversed direction in the superior temporal areas in logographic language compared to earlier reports on alphabetic language (Raij et al., 2000; van Atteveldt et al., 2004). Converging evidence from another audiovisual integration study (Jost et al., 2014) revealed a similar congruency effect for familiar German words, and was attributed to lexical-semantic processes involved in the processing of audiovisual words stimuli. The inferior frontal cortex has repeatedly been shown to be activated specifically by semantically incongruent audiovisual stimuli, which has been attributed to increased demands on the cognitive control involving semantic retrieval and working memory processes (Martin and Chao, 2001; Wagner et al., 2001; Doehrmann and Naumer, 2008).

The Chinese group showed a decrease of activation in the inferior frontal cortex when congruent audiovisual pairs were compared to the combined auditory and visual only stimulations. This indicates less neural resources for the audiovisual processing than the sum of resources it took to process the auditory and visual information independently. Furthermore, in the congruency comparison, the incongruent audiovisual pairs showed a higher activation than the congruent pairs again suggesting that it is less demanding to process the congruent audiovisual information due to complementary representations from both auditory and visual modalities. Since the Finnish participants had never learned the character-speech sound associations, they did not show any congruency effects. The left-lateralized suppression and congruency effects in the Chinese group suggest that the native Chinese could effectively utilize the audiovisual features of the learned language, whereas the right-lateralized suppression effect in the Finnish group may suggest that the Finnish participants rely more on the basic audiovisual processing mechanisms including substantial visual feature analysis. Therefore, congruency and suppression effects provide complementary information about audiovisual integration.

Both suppression and congruency effects in the Chinese group were mainly in the late time window about 500–800 ms after

the stimulus onset. Previous studies have also found similar late orthographic-phonological interactions (400–700 ms) in spoken language processing during metaphonological tasks (Pattamadilok et al., 2011; Lafontaine et al., 2012). Interestingly, a late negativity enhancement (about 650 ms) after auditory stimulus onset was reported in mismatch negativity (MMN) studies on audiovisual integration in both the beginner and advanced readers in children (Froyen et al., 2009; Zaric et al., 2014, 2015). It was interpreted as more elaborate associative processes that are activated for the integration of letter-speech pairs in children. In addition, the suppression and congruency effects were mainly localized in the left frontal and superior and middle temporal areas. These locations are consistent with previous fMRI studies using Chinese characters and have shown that character reading process is characterized by particularly strong left lateralization of the frontal (BAs 9 and 47) and temporal cortices (Tan et al., 2000, 2001). For example, the left middle frontal area (BA 9) was suggested as an important region for integrating the logographic visuospatial analysis and the semantic processing in Chinese (Tan et al., 2001). Given the fact that the characters/speech items all refer to meaningful words, both the suppression as well as the congruency effect in the late time window most likely reflect a mixture of modulatory feedback and lexico-semantic processing during audiovisual integration in the Chinese participants. However, as discussed earlier, such audiovisual modulatory feedback and lexico-semantic processing seem to be largely overlapping in time windows and brain regions, further studies are needed to dissociate the effects of semantic processing from the other audiovisual process.

There are some limitations in the present study. We used an active paradigm which corresponded to real life situations of audiovisual integration (e.g., learning to read) as an active action. However, it also brought some challenges to identify the underlying functional brain networks while the participants were performing a working memory task. The brain activity during an active task makes the interpretation of direct group comparisons problematic since the task difficulties and demands for the two groups were different. Therefore, the conclusion regarding the orthographic influences on multisensory integration was indirectly based on the comparison of our results in the Chinese group with previous letter-speech sound integration studies in alphabetic languages. Another limitation relates to the lack of structural magnetic resonance images in the present study, which could potentially lead to poorer source localization accuracy. However, the localization of the brain activity at the millimeter scale was not the goal of the current study, instead we were interested in the estimation of the larger cortical areas that contributed to the cognitive process.

CONCLUSION

Our findings demonstrated the effect of long-term exposure to logographic language on audiovisual integration processes

for written characters and speech sounds. Different suppression effects in the Chinese and Finnish groups indicated the adaptive nature of the brain networks for processing different types of audiovisual information. Importantly, in the Chinese group, the left superior temporal and inferior frontal cortices were actively involved in the processing of both audiovisual vs. unimodal and congruency information pinpointing the left superior temporal and frontal cortex as important hubs for audiovisual and semantic processing in Chinese. The findings are remarkably similar to those found for alphabetic languages in the left superior temporal cortex but with also unique aspects of the processing of the logographic stimuli in the inferior frontal cortex. The results thus indicate that there are universal audio-visual association mechanisms in language learning complemented by language-specific processes.

REFERENCES

- Ahveninen, J., Jääskeläinen, I. P., Raij, T., Bonmassar, G., Devore, S., Hämäläinen, M., et al. (2006). Task-modulated “what” and “where” pathways in human auditory cortex. *Proc. Natl. Acad. Sci. U S A* 103, 14608–14613. doi: 10.1073/pnas.0510480103
- Andersen, T. S., Tiippana, K., and Sams, M. (2004). Factors influencing audiovisual fission and fusion illusions. *Cogn. Brain Res.* 21, 301–308. doi: 10.1016/j.cogbrainres.2004.06.004
- Bernstein, L. E., and Liebenthal, E. (2014). Neural pathways for visual speech perception. *Front. Neurosci.* 8:386. doi: 10.3389/fnins.2014.00386
- Besle, J., Fort, A., and Giard, M.-H. (2004). Interest and validity of the additive model in electrophysiological studies of multisensory interactions. *Cogn. Process.* 5, 189–192. doi: 10.1007/s10339-004-0026-y
- Blau, V., van Atteveldt, N., Formisano, E., Goebel, R., and Blomert, L. (2008). Task-irrelevant visual letters interact with the processing of speech sounds in heteromodal and unimodal cortex. *Eur. J. Neurosci.* 28, 500–509. doi: 10.1111/j.1460-9568.2008.06350.x
- Brem, S., Bach, S., Kucian, K., Kujala, J. V., Guttorm, T. K., Martin, E., et al. (2010). Brain sensitivity to print emerges when children learn letter-speech sound correspondences. *Proc. Natl. Acad. Sci. U S A* 107, 7939–7944. doi: 10.1073/pnas.0904402107
- Bremmer, F., Schlack, A., Shah, N. J., Zafiris, O., Kubischik, M., Hoffmann, K., et al. (2001). Polymodal motion processing in posterior parietal and premotor cortex: a human fMRI study strongly implies equivalencies between humans and monkeys. *Neuron* 29, 287–296. doi: 10.1016/s0896-6273(01)00198-2
- Calvert, G. A. (2001). Crossmodal processing in the human brain: insights from functional neuroimaging studies. *Cereb. Cortex* 11, 1110–1123. doi: 10.1093/cercor/11.12.1110
- Calvert, G. A., Hansen, P. C., Iversen, S. D., and Brammer, M. J. (2001). Detection of audio-visual integration sites in humans by application of electrophysiological criteria to the BOLD effect. *Neuroimage* 14, 427–438. doi: 10.1006/nimg.2001.0812
- Calvert, G. A., and Thesen, T. (2004). Multisensory integration: methodological approaches and emerging principles in the human brain. *J. Physiol. Paris* 98, 191–205. doi: 10.1016/j.jphysparis.2004.03.018
- Campanella, S., and Belin, P. (2007). Integrating face and voice in person perception. *Trends Cogn. Sci.* 11, 535–543. doi: 10.1016/j.tics.2007.10.001
- Campbell, R. (2008). The processing of audio-visual speech: empirical and neural bases. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 363, 1001–1010. doi: 10.1098/rstb.2007.2155
- Cappe, C., Thut, G., Romei, V., and Murray, M. M. (2010). Auditory-visual multisensory interactions in humans: timing, topography, directionality, and sources. *J. Neurosci.* 30, 12572–12580. doi: 10.1523/JNEUROSCI.1099-10.2010
- Cohen, Y. E. (2009). Multimodal activity in the parietal cortex. *Hear. Res.* 258, 100–105. doi: 10.1016/j.heares.2009.01.011
- Dale, A. M., Liu, A. K., Fischl, B. R., Buckner, R. L., Belliveau, J. W., Lewine, J. D., et al. (2000). Dynamic statistical parametric mapping. *Neuron* 26, 55–67. doi: 10.1016/S0896-6273(00)81138-1
- Doehrmann, O., and Naumer, M. J. (2008). Semantics and the multisensory brain: how meaning modulates processes of audio-visual integration. *Brain Res.* 1242, 136–150. doi: 10.1016/j.brainres.2008.03.071
- Du, Y., Zhang, Q., and Zhang, J. X. (2014). Does N200 reflect semantic processing?—an ERP study on chinese visual word recognition. *PLoS One* 9:e90794. doi: 10.1371/journal.pone.0090794
- Fort, A., Delpuech, C., Pernier, J., and Giard, M. H. (2002). Dynamics of cortico-subcortical cross-modal operations involved in audio-visual object detection in humans. *Cereb. Cortex* 12, 1031–1039. doi: 10.1093/cercor/12.10.1031
- Foxe, J. J., Morocz, I. A., Murray, M. M., Higgins, B. A., Javitt, D. C., and Schroeder, C. E. (2000). Multisensory auditory-somatosensory interactions in early cortical processing revealed by high-density electrical mapping. *Cogn. Brain Res.* 10, 77–83. doi: 10.1016/s0926-6410(00)00024-0
- Froyen, D. J. W., Bonte, M. L., van Atteveldt, N., and Blomert, L. (2009). The long road to automation: neurocognitive development of letter-speech sound processing. *J. Cogn. Neurosci.* 21, 567–580. doi: 10.1162/jocn.2009.21061
- Gramfort, A., Luessi, M., Larson, E., Engemann, D. A., Strohmeier, D., Brodbeck, C., et al. (2013). MEG and EEG data analysis with MNE-Python. *Front. Neurosci.* 7:267. doi: 10.3389/fnins.2013.00267
- Grunewald, A., Linden, J. F., and Andersen, R. A. (1999). Responses to auditory stimuli in macaque lateral intraparietal area I: effects of training. *J. Neurophysiol.* 82, 330–342. doi: 10.1152/jn.1999.82.1.330
- Hämäläinen, M. S., and Ilmoniemi, R. J. (1994). Interpreting magnetic fields of the brain: minimum norm estimates. *Med. Biol. Eng. Comput.* 32, 35–42. doi: 10.1007/bf02512476
- Hein, G., Doehrmann, O., Müller, N. G., Kaiser, J., Muckli, L., and Naumer, M. J. (2007). Object familiarity and semantic congruency modulate responses in cortical audiovisual integration areas. *J. Neurosci.* 27, 7881–7887. doi: 10.1523/JNEUROSCI.1740-07.2007
- Herdman, A. T., Fujioka, T., Chau, W., Ross, B., Pantev, C., and Picton, T. W. (2006). Cortical oscillations related to processing congruent and incongruent grapheme-phoneme pairs. *Neurosci. Lett.* 399, 61–66. doi: 10.1016/j.neulet.2006.01.069
- Hocking, J., and Price, C. J. (2009). Dissociating verbal and nonverbal audiovisual object processing. *Brain Lang.* 108, 89–96. doi: 10.1016/j.bandl.2008.10.005
- Holloway, I. D., van Atteveldt, N., Blomert, L., and Ansari, D. (2015). Orthographic dependency in the neural correlates of reading: evidence from audiovisual integration in English readers. *Cereb. Cortex* 25, 1544–1553. doi: 10.1093/cercor/bht347

AUTHOR CONTRIBUTIONS

WX, JH, and OK designed the study and performed the MEG experiments. WX, JH, and RO analyzed the data. All authors discussed the results and contributed to the final manuscript.

FUNDING

This study was supported by the European Union H2020 Marie Skłodowska-Curie Actions (MSCA)-ITN-2014-ETN Programme, “Advancing brain research in children’s developmental neurocognitive disorders” project (ChildBrain, #641652), “Understanding and predicting developmental language abilities and disorders in multilingual Europe” project (Predictable, #641858) and the Academy of Finland (MultiLeTe, #292466).

- Jones, J. A., and Callan, D. E. (2003). Brain activity during audiovisual speech perception: an fMRI study of the McGurk effect. *Neuroreport* 14, 1129–1133. doi: 10.1097/00001756-200306110-00006
- Jost, L. B., Eberhard-Moscicka, A. K., Frisch, C., Dellwo, V., and Maurer, U. (2014). Integration of spoken and written words in beginning readers: a topographic ERP study. *Brain Topogr.* 27, 786–800. doi: 10.1007/s10548-013-0336-4
- Kamke, M. R., Vieth, H. E., Cottrell, D., and Mattingley, J. B. (2012). Parietal disruption alters audiovisual binding in the sound-induced flash illusion. *Neuroimage* 62, 1334–1341. doi: 10.1016/j.neuroimage.2012.05.063
- Kayser, C., Petkov, C. I., and Logothetis, N. K. (2008). Visual modulation of neurons in auditory cortex. *Cereb. Cortex* 18, 1560–1574. doi: 10.1093/cercor/bhm187
- Kayser, S. J., Philiastides, M. G., and Kayser, C. (2017). Sounds facilitate visual motion discrimination via the enhancement of late occipital visual representations. *Neuroimage* 148, 31–41. doi: 10.1016/j.neuroimage.2017.01.010
- Kuo, W.-J., Yeh, T.-C., Duann, J.-R., Wu, Y.-T., Ho, L.-T., Hung, D., et al. (2001). A left-lateralized network for reading Chinese words: a 3 T fMRI study. *Neuroreport* 12, 3997–4001. doi: 10.1097/00001756-200112210-00029
- Kutas, M., and Federmeier, K. D. (2011). Thirty years and counting: finding meaning in the N400 component of the event-related brain potential (ERP). *Annu. Rev. Psychol.* 62, 621–647. doi: 10.1146/annurev.psych.093008.131123
- Lafontaine, H., Chetail, F., Colin, C., Kolinsky, R., and Pattamadilok, C. (2012). Role and activation time course of phonological and orthographic information during phoneme judgments. *Neuropsychologia* 50, 2897–2906. doi: 10.1016/j.neuropsychologia.2012.08.020
- Laurienti, P. J., Perrault, T. J., Stanford, T. R., Wallace, M. T., and Stein, B. E. (2005). On the use of superadditivity as a metric for characterizing multisensory integration in functional neuroimaging studies. *Exp. Brain Res.* 166, 289–297. doi: 10.1007/s00221-005-2370-2
- Lewis, J. W., Beauchamp, M. S., and DeYoe, E. A. (2000). A comparison of visual and auditory motion processing in human cerebral cortex. *Cereb. Cortex* 10, 873–888. doi: 10.1093/cercor/10.9.873
- Lieberman, A. M. (1992). Chapter 9 the relation of speech to reading and writing. *Adv. Psychol.* 94, 167–178. doi: 10.1016/s0166-4115(08)62794-6
- Lin, F. H., Witzel, T., Ahlfors, S. P., Stufflebeam, S. M., Belliveau, J. W., and Hämäläinen, M. S. (2006). Assessing and improving the spatial accuracy in MEG source localization by depth-weighted minimum-norm estimates. *Neuroimage* 31, 160–171. doi: 10.1016/j.neuroimage.2005.11.054
- Lütkenhöner, B., Lammertmann, C., Simões, C., and Hari, R. (2002). Magnetoencephalographic correlates of audiotactile interaction. *Neuroimage* 15, 509–522. doi: 10.1006/nimg.2001.0991
- Maded, S., Le Goff, K., Anton, J.-L., Longcamp, M., Velay, J.-L., Nazarian, B., et al. (2016). Brain correlates of phonological recoding of visual symbols. *Neuroimage* 132, 359–372. doi: 10.1016/j.neuroimage.2016.02.010
- Maris, E., and Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *J. Neurosci. Methods* 164, 177–190. doi: 10.1016/j.jneumeth.2007.03.024
- Martin, A., and Chao, L. L. (2001). Semantic memory and the brain: structure and processes. *Curr. Opin. Neurobiol.* 11, 194–201. doi: 10.1016/s0959-4388(00)00196-3
- Maurer, U., Brem, S., Kranz, F., Bucher, K., Benz, R., Halder, P., et al. (2006). Coarse neural tuning for print peaks when children learn to read. *Neuroimage* 33, 749–758. doi: 10.1016/j.neuroimage.2006.06.025
- McBride, C. A. (2015). Is chinese special? Four aspects of chinese literacy acquisition that might distinguish learning chinese from learning alphabetic orthographies. *Educ. Psychol. Rev.* 28, 523–549. doi: 10.1007/s10648-015-9318-2
- Meredith, M. A. (2002). On the neuronal basis for multisensory convergence: a brief overview. *Cogn. Brain Res.* 14, 31–40. doi: 10.1016/s0926-6410(02)00059-9
- Molholm, S., Ritter, W., Murray, M. M., Javitt, D. C., Schroeder, C. E., and Foxe, J. J. (2002). Multisensory auditory-visual interactions during early sensory processing in humans: a high-density electrical mapping study. *Cogn. Brain Res.* 14, 115–128. doi: 10.1016/s0926-6410(02)00066-6
- Murray, M. M., Lewkowicz, D. J., Amedi, A., and Wallace, M. T. (2016a). Multisensory processes: a balancing act across the lifespan. *Trends Neurosci.* 39, 567–579. doi: 10.1016/j.tins.2016.05.003
- Murray, M. M., Thelen, A., Thut, G., Romei, V., Martuzzi, R., and Matusz, P. J. (2016b). The multisensory function of the human primary visual cortex. *Neuropsychologia* 83, 161–169. doi: 10.1016/j.neuropsychologia.2015.08.011
- Murray, M. M., and Spierer, L. (2009). Auditory spatio-temporal brain dynamics and their consequences for multisensory interactions in humans. *Hear. Res.* 258, 121–133. doi: 10.1016/j.heares.2009.04.022
- Ojanen, V., Möttönen, R., Pekkola, J., Jääskeläinen, I. P., Joensuu, R., Autti, T., et al. (2005). Processing of audiovisual speech in Broca's area. *Neuroimage* 25, 333–338. doi: 10.1016/j.neuroimage.2004.12.001
- Oostenveld, R., Fries, P., Maris, E., and Schoffelen, J.-M. (2011). FieldTrip: open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Comput. Intell. Neurosci.* 2011:156869. doi: 10.1155/2011/156869
- Pattamadilok, C., Perre, L., and Ziegler, J. C. (2011). Beyond rhyme or reason: ERPs reveal task-specific activation of orthography on spoken language. *Brain Lang.* 116, 116–124. doi: 10.1016/j.bandl.2010.12.002
- Perfetti, C. A., and Tan, L. H. (1998). The time course of graphic, phonological, and semantic activation in Chinese character identification. *J. Exp. Psychol. Learn. Mem. Cogn.* 24, 101–118. doi: 10.1037/0278-7393.24.1.101
- Perrault, T. J. Jr., Vaughan, J. W., Stein, B. E., and Wallace, M. T. (2005). Superior colliculus neurons use distinct operational modes in the integration of multisensory stimuli. *J. Neurophysiol.* 93, 2575–2586. doi: 10.1152/jn.00926.2004
- Price, C. J. (2000). The anatomy of language: contributions from functional neuroimaging. *J. Anat.* 197, 335–359. doi: 10.1046/j.1469-7580.2000.1973.0335.x
- Pugh, K. R., Mencl, W. E., Shaywitz, B. A., Shaywitz, S. E., Fulbright, R. K., Constable, R. T., et al. (2000). The angular gyrus in developmental dyslexia: task-specific differences in functional connectivity within posterior cortex. *Psychol. Sci.* 11, 51–56. doi: 10.1111/1467-9280.00214
- Raij, T., Uutela, K., and Hari, R. (2000). Audiovisual integration of letters in the human brain. *Neuron* 28, 617–625. doi: 10.1016/s0896-6273(00)00138-0
- Rüsseler, J., Ye, Z., Gerth, I., Szycik, G. R., and Münte, T. F. (2018). Audio-visual speech perception in adult readers with dyslexia: an fMRI study. *Brain Imaging Behav.* 12, 357–368. doi: 10.1007/s11682-017-9694-y
- Schlaggar, B. L., and McCandliss, B. D. (2007). Development of neural systems for reading. *Annu. Rev. Neurosci.* 30, 475–503. doi: 10.1146/annurev.neuro.28.061604.135645
- Schroeder, C. E., and Foxe, J. J. (2002). The timing and laminar profile of converging inputs to multisensory areas of the macaque neocortex. *Cogn. Brain Res.* 14, 187–198. doi: 10.1016/s0926-6410(02)00073-3
- Schröger, E., and Widmann, A. (1998). Speeded responses to audiovisual signal changes result from bimodal integration. *Psychophysiology* 35, 755–759. doi: 10.1111/1469-8986.3560755
- Service, E., Helenius, P., Maury, S., and Salmelin, R. (2007). Localization of syntactic and semantic brain responses using magnetoencephalography. *J. Cogn. Neurosci.* 19, 1193–1205. doi: 10.1162/jocn.2007.19.7.1193
- Shu, H. (2003). Chinese writing system and learning to read. *Int. J. Psychol.* 38, 274–285. doi: 10.1080/00207590344000060
- Simos, P. G., Rezaie, R., Fletcher, J. M., and Papanicolaou, A. C. (2013). Time-constrained functional connectivity analysis of cortical networks underlying phonological decoding in typically developing school-aged children: a magnetoencephalography study. *Brain Lang.* 125, 156–164. doi: 10.1016/j.bandl.2012.07.003
- Sperdin, H. F., Cappe, C., Foxe, J. J., and Murray, M. M. (2009). Early, low-level auditory-somatosensory multisensory interactions impact reaction time speed. *Front. Integr. Neurosci.* 3:2. doi: 10.3389/fnint.2009.07.002.2009
- Stein, B. E., and Stanford, T. R. (2008). Multisensory integration: current issues from the perspective of the single neuron. *Nat. Rev. Neurosci.* 9, 255–266. doi: 10.1038/nrn2331
- Tan, L. H., Liu, H. L., Perfetti, C. A., Spinks, J. A., Fox, P. T., and Gao, J. H. (2001). The neural system underlying Chinese logograph reading. *Neuroimage* 13, 836–846. doi: 10.1006/nimg.2001.0749

- Tan, L. H., Spinks, J. A., Eden, G. F., Perfetti, C. A., and Siok, W. T. (2005). Reading depends on writing, in Chinese. *Proc. Natl. Acad. Sci. U S A* 102, 8781–8785. doi: 10.1073/pnas.0503523102
- Tan, L. H., Spinks, J. A., Gao, J.-H., Liu, H.-L., Perfetti, C. A., Xiong, J., et al. (2000). Brain activation in the processing of Chinese characters and words: a functional MRI study. *Hum. Brain Mapp.* 10, 16–27. doi: 10.1002/(sici)1097-0193(200005)10:1<16::aid-hbm30>3.0.co;2-m
- Taulu, S., and Kajola, M. (2005). Presentation of electromagnetic multichannel data: the signal space separation method. *J. Appl. Phys.* 97:124905. doi: 10.1063/1.1935742
- Taulu, S., Kajola, M., and Simola, J. (2004). Suppression of interference and artifacts by the signal space separation method. *Brain Topogr.* 16, 269–275. doi: 10.1023/b:brat.0000032864.93890.f9
- Taulu, S., and Simola, J. (2006). Spatiotemporal signal space separation method for rejecting nearby interference in MEG measurements. *Phys. Med. Biol.* 51, 1759–1768. doi: 10.1088/0031-9155/51/7/008
- Teder-Sälejärvi, W. A., McDonald, J. J., Di Russo, F., and Hillyard, S. A. (2002). An analysis of audio-visual crossmodal integration by means of event-related potential (ERP) recordings. *Cogn. Brain Res.* 14, 106–114. doi: 10.1016/s0926-6410(02)00065-4
- van Atteveldt, N., and Ansari, D. (2014). How symbols transform brain function: a review in memory of Leo Blomert. *Trends Neurosci. Edu.* 3, 44–49. doi: 10.1016/j.tine.2014.04.001
- van Atteveldt, N. M., Formisano, E., Blomert, L., and Goebel, R. (2007a). The effect of temporal asynchrony on the multisensory integration of letters and speech sounds. *Cereb. Cortex* 17, 962–974. doi: 10.1093/cercor/bhl007
- van Atteveldt, N. M., Formisano, E., Goebel, R., and Blomert, L. (2007b). Top-down task effects overrule automatic multisensory responses to letter-sound pairs in auditory association cortex. *Neuroimage* 36, 1345–1360. doi: 10.1016/j.neuroimage.2007.03.065
- van Atteveldt, N., Formisano, E., Goebel, R., and Blomert, L. (2004). Integration of letters and speech sounds in the human brain. *Neuron* 43, 271–282. doi: 10.1016/j.neuron.2004.06.025
- van Atteveldt, N., Murray, M. M., Thut, G., and Schroeder, C. E. (2014). Multisensory integration: flexible use of general operations. *Neuron* 81, 1240–1253. doi: 10.1016/j.neuron.2014.02.044
- van Atteveldt, N., Roebroek, A., and Goebel, R. (2009). Interaction of speech and script in human auditory cortex: insights from neuro-imaging and effective connectivity. *Hear. Res.* 258, 152–164. doi: 10.1016/j.heares.2009.05.007
- Vartiainen, J., Parviainen, T., and Salmelin, R. (2009). Spatiotemporal convergence of semantic processing in reading and speech perception. *J. Neurosci.* 29, 9271–9280. doi: 10.1523/JNEUROSCI.5860-08.2009
- Wagner, A. D., Paré-Blagoev, E. J., Clark, J., and Poldrack, R. A. (2001). Recovering meaning: left prefrontal cortex guides controlled semantic retrieval. *Neuron* 31, 329–338. doi: 10.1016/S0896-6273(01)00359-2
- Wang, S., Zhu, Z., Zhang, J. X., Wang, Z., Xiao, Z., Xiang, H., et al. (2008). Broca's area plays a role in syntactic processing during Chinese reading comprehension. *Neuropsychologia* 46, 1371–1378. doi: 10.1016/j.neuropsychologia.2007.12.020
- Wu, C.-Y., Ho, M.-H. R., and Chen, S.-H. A. (2012). A meta-analysis of fMRI studies on Chinese orthographic, phonological, and semantic processing. *Neuroimage* 63, 381–391. doi: 10.1016/j.neuroimage.2012.06.047
- Xu, W., Kolozsvári, O. B., Monto, S. P., and Hämäläinen, J. A. (2018). Brain responses to letters and speech sounds and their correlations with cognitive skills related to reading in children. *Front. Hum. Neurosci.* 12:304. doi: 10.3389/fnhum.2018.00304
- Žarić, G., Fraga González, G., Tijms, J., van der Molen, M. W., Blomert, L., and Bonte, M. (2014). Reduced neural integration of letters and speech sounds in dyslexic children scales with individual differences in reading fluency. *PLoS One* 9:e110337. doi: 10.1371/journal.pone.0110337
- Žarić, G., Fraga González, G., Tijms, J., van der Molen, M. W., Blomert, L., and Bonte, M. (2015). Crossmodal deficit in dyslexic children: practice affects the neural timing of letter-speech sound integration. *Front. Hum. Neurosci.* 9:369. doi: 10.3389/fnhum.2015.00369
- Ziegler, J. C., and Goswami, U. (2005). Reading acquisition, developmental dyslexia, and skilled reading across languages: a psycholinguistic grain size theory. *Psychol. Bull.* 131, 3–29. doi: 10.1037/0033-2909.131.1.3

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Xu, Kolozsvári, Oostenveld, Leppänen and Hämäläinen. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



II

BRAIN RESPONSES TO LETTERS AND SPEECH SOUNDS AND THEIR CORRELATIONS WITH COGNITIVE SKILLS RELATED TO READING IN CHILDREN

by

Weiyong Xu, Orsolya B. Kolozsvári, Simo P. Monto and Jarmo A. Hämäläinen,
2018

Frontiers in Human Neuroscience, 12, 304

<https://doi.org/10.3389/fnhum.2018.00304>

Reproduced with kind permission by Frontiers.



Brain Responses to Letters and Speech Sounds and Their Correlations With Cognitive Skills Related to Reading in Children

Weiyong Xu^{1,2*}, Orsolya B. Kolozsvári^{1,2}, Simo P. Monto^{1,2} and Jarmo A. Hämäläinen^{1,2}

¹ Department of Psychology, University of Jyväskylä, Jyväskylä, Finland, ² Jyväskylä Centre for Interdisciplinary Brain Research, University of Jyväskylä, Jyväskylä, Finland

Letter-speech sound (LSS) integration is crucial for initial stages of reading acquisition. However, the relationship between cortical organization for supporting LSS integration, including unimodal and multimodal processes, and reading skills in early readers remains unclear. In the present study, we measured brain responses to Finnish letters and speech sounds from 29 typically developing Finnish children in a child-friendly audiovisual integration experiment using magnetoencephalography. Brain source activations in response to auditory, visual and audiovisual stimuli as well as audiovisual integration response were correlated with reading skills and cognitive skills predictive of reading development after controlling for the effect of age. Regression analysis showed that from the brain measures, the auditory late response around 400 ms showed the largest association with phonological processing and rapid automatized naming abilities. In addition, audiovisual integration effect was most pronounced in the left and right temporoparietal regions and the activities in several of these temporoparietal regions correlated with reading and writing skills. Our findings indicated the important role of temporoparietal regions in the early phase of learning to read and their unique contribution to reading skills.

Keywords: letter-speech sound integration, brain development, magnetoencephalography, auditory cortex, language learning, reading

OPEN ACCESS

Edited by:

Gorka Fraga González,
University of Amsterdam, Netherlands

Reviewed by:

Joao Correia,
Maastricht University, Netherlands
Mario Braun,
University of Salzburg, Austria

*Correspondence:

Weiyong Xu
weiyong.w.xu@jyu.fi

Received: 30 April 2018

Accepted: 16 July 2018

Published: 03 August 2018

Citation:

Xu W, Kolozsvári OB, Monto SP and Hämäläinen JA (2018) Brain Responses to Letters and Speech Sounds and Their Correlations With Cognitive Skills Related to Reading in Children.
Front. Hum. Neurosci. 12:304.
doi: 10.3389/fnhum.2018.00304

INTRODUCTION

Letter-speech sound (LSS) integration is a key step in learning to read for alphabetic languages. The development and reorganization of early readers' language circuits for supporting automatized LSS integration and how such integration is related to the development of fluent reading are crucial questions from both theoretical and practical point of view (Shankweiler et al., 2008; Dehaene et al., 2015). Research has shown that in early readers, the print-speech convergence (as measured by coactivation in fMRI) in the left reading network (inferior frontal gyrus, inferior parietal cortex, and fusiform gyrus) is a significant predictor of reading achievement measured 2 years later (Preston et al., 2016). In another study using four contrasting languages to find common indices of successful literacy acquisition, highly similar neural organization for print-speech convergence was observed between the languages. Furthermore, such print-speech convergence was suggested as a common brain signature of reading proficiency (Rueckl et al., 2015). However, little is known

about the interrelationships between brain mechanisms of speech perception, letter processing, LSS integration and the development of reading skills during childhood.

In order to understand the development of LSS integration, which is a form of audiovisual integration, auditory and visual processes also need to be taken into account. The maturation of auditory and visual cortices is reflected by changes in the auditory and visual evoked responses. In general, the auditory evoked responses have been shown to change greatly with the tendency of shortened latencies and decreased amplitudes from childhood to adulthood (Albrecht et al., 2000). For example, the auditory P1 and N1b (the supratemporal component of the N1) peaks show large age-related decreases in latency. In addition, auditory P1, P1-N1b, and N2 peak amplitudes change throughout childhood with accelerated change around the age of 10 years (Ponton et al., 2000). For the visual components, there is a clear delay in the activation timing in children compared to adults, which progressively increases from occipital (related to low-level visual analysis) to occipitotemporal (related to letters/letter string analysis) and further to temporal areas (related to written word perception) (Parviainen et al., 2006).

It has been shown that audiovisual speech produces audiovisual interaction effects reflected as both suppression of the visual response to lipreading and reduced auditory responses to the speech sound compared with unimodal conditions (Besle et al., 2004, 2008). One study used audiovisual speech and audiovisual non-linguistic stimuli to investigate the developmental pattern of audiovisual interactions in the age range of 5–19 years (Tremblay et al., 2007). The results showed that the strength of audiovisual speech integration significantly correlated with age, whereas the performance on non-speech tasks seemed to be similar across all ages. These findings suggest independent maturational processes for audiovisual speech and non-speech during childhood. Converging evidence from electrophysiological research revealed a systematic relationship between brain responses underlying audiovisual integration (of simple audiovisual sounds and objects) in the time range of the auditory N1 response (about 120 ms) and age between 7 and 16 years (Brandwein et al., 2011). Multisensory processes are thus still developing even in late childhood and this maturation is likely to be reflected in learning and automatization of LSS correspondences, as well as in the associations with reading skill development.

As children learn to read, their sensitivities to print are paralleled by changes in an occipitotemporal negativity N1 (or N170) to words as measured by event-related potentials (Brem et al., 2010; Maurer et al., 2010; Fraga González et al., 2014). This visual N1 has been shown to develop with reading skills, showing an inverted U-shaped developmental trajectory with maximum N1 tuning (selectivity for print) during the second grade and further decrease of the N1 tuning in adults (Maurer et al., 2005, 2006). Neuroimaging studies have localized the visual print-sensitive N1 in a region within the left fusiform gyrus called “visual word form area” (VWFA) (McCandliss et al., 2003; Dehaene and Cohen, 2011). In one recent study (Bach et al., 2013) 19 non-reading kindergarteners were trained in letter-speech sound associations with Graphogame (Lyytinen

et al., 2009) for 8 weeks. It was found that the N1 and the VWFA activation in these kindergarteners significantly improved the prediction of reading skills in second grade over behavioral data alone and together with the behavioral measures they explained up to 88% of the variance in reading (Bach et al., 2013). Therefore, visual N1 is considered as a sensitive index of visual letter string processing reflecting important processes for reading fluency (Fraga González et al., 2014, 2017).

Audiovisual integration, which is defined as the interaction between auditory and visual modalities, and its developmental trajectory remain poorly understood. The additive model, which is based on the comparison of multisensory responses to the summed responses from the constituent unisensory conditions [responses to audiovisual stimuli – responses to (auditory stimuli + visual stimuli)], has been frequently used in electrophysiological studies on multisensory integration (Calvert and Thesen, 2004; Stein and Stanford, 2008; Sperdin et al., 2009). Another commonly used approach in audiovisual research is to study the congruency effect (Jones and Callan, 2003; Ojanen et al., 2005; Hein et al., 2007; Rüsseler et al., 2017), which involves a contrast between congruent and incongruent audiovisual pairs. LSS in alphabetic languages consistently activates several language and cross-modal brain regions in adults. Regions particularly in the superior temporal cortices, have been shown in fMRI studies to have heteromodal properties (van Atteveldt et al., 2009). These brain regions have also been implicated in magnetoencephalography (MEG) findings showing LSS sites in the left and right superior temporal sulci (STS) (Raij et al., 2000). Feedback projections from this heteromodal region have also been shown in fMRI studies to modify the response in a modality-specific region of the primary auditory cortex (van Atteveldt et al., 2004). Top-down factors generated by different task demands and instructions also clearly impact multisensory integration (Andersen et al., 2004). For example, use of explicit vs. implicit and passive vs. active experimental task has been shown to influence the brain responses related to LSS (van Atteveldt et al., 2007; Blau et al., 2008).

Accessing the phonological representations for written words and letter strings has been shown to also involve the parietal areas in many studies particularly the supramarginal gyrus (BA 40) and the angular gyrus (BA 39) (Price, 2000; Pugh et al., 2000b; Schlaggar and McCandliss, 2007). Activation in these two posterior regions was found to significantly correlate with cross-modal (auditory and visual) language task performance (Booth et al., 2003). Furthermore, neuroimaging studies have confirmed that activation in the angular gyrus and supramarginal gyrus were associated with phonological (Buchsbaum and D’Esposito, 2008; Sliwinska et al., 2015) and semantic processing (Binder et al., 2009) of written words, respectively. Parietal regions also show differences during phonological processing in children with reading difficulties (Vandermosten et al., 2016). Taken together, there are several temporoparietal brain regions that are suggested to be involved in the process of integrating visual and auditory information for the purpose of reading.

In contrast to the natural relationship between auditory and visual information in audiovisual speech, the association between letters and speech sounds is mostly based upon agreed conventions. Although knowledge of letter-speech sound associations seems easy to acquire within 1 year of reading instruction (Hardy et al., 1972), EEG studies using mismatch negativity (MMN) paradigm (Näätänen, 2001) have found that beginning readers showed protracted development of letter-speech sound associations beyond early school years (Froyen et al., 2009) and such orthographic-phonological integration could serve as a neural signature of successful or failing reading development (Blomert, 2011). Studies on dyslexia have revealed reduced audiovisual integration (indexed by cross modal MMN) which is associated with a more fundamental deficit in the auditory processing of speech sounds leading to reading failure (Blau et al., 2009; Žarić et al., 2014). Therefore, audiovisual integration is considered as an important marker associated with reading fluency and has been shown to facilitate visual specialization (indexed by print sensitive N1 in the VWFA) in learning to read (Fraga González et al., 2016, 2017).

Although LSS integration has been shown to be important for reading development (Blau et al., 2009, 2010; Blomert and Froyen, 2010; Blomert, 2011), reading is also dependent on other cognitive skills. Several behavioral measures such as phonological awareness, verbal short-term memory and rapid automatized naming (RAN) have been shown to be closely associated with reading skills and provide a good estimation of risk for dyslexia (Pennington and Lefly, 2001; Puolakanaho et al., 2007; Melby-Lervåg et al., 2012). These cognitive measures have been shown to be important mediators of the prediction of reading outcome from brain responses as measured by ERPs (Lohvansuu et al., 2018).

In the present study, we measured auditory responses to speech and visual responses to letters as well as audiovisual integration related responses of letter-speech sound combinations with MEG with the purpose of linking these brain responses to reading development. Previous studies (Froyen et al., 2008, 2009; Blomert, 2011) have often used an audiovisual oddball design and shown a long developmental trajectory for LSS integration. We used an experimental design with equal numbers of unimodal and bimodal stimuli as well as equal numbers of congruent and incongruent audiovisual stimuli. This allows a more direct examination of the LSS integration as well as separating the unimodal effects from the audiovisual effects. We used regression-based methods (controlling for age) to explore the relationship between the neural-level responses to speech sounds, visual letters, audiovisual combinations and behavioral cognitive skills. We expected to see associations between responses to the speech sounds and phonological and reading skills (e.g., Lohvansuu et al., 2018), between the visual N1 and reading skills (e.g., Brem et al., 2010; Maurer et al., 2010; Fraga González et al., 2014), and importantly between the brain measures of LSS integration and reading skills (Blau et al., 2009; Blomert and Willems, 2010; Blomert, 2011; Preston et al., 2016; Fraga González et al., 2017).

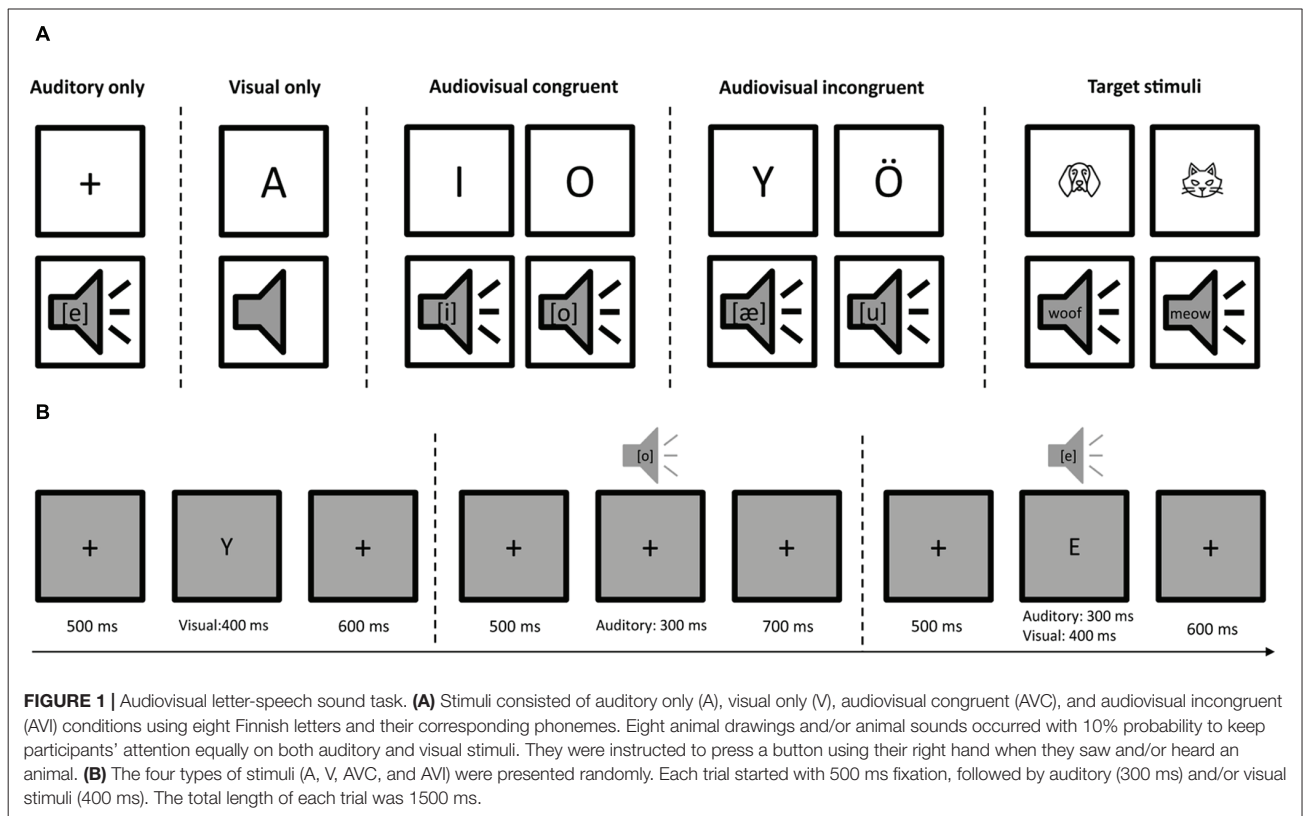
MATERIALS AND METHODS

Participants

All participants were Finnish speaking school children (6–11 years) recruited through the National Registry of Finland. None of the participants had neurological disorders or problems caused by permanent head injuries, ADHD, delay in language development or language-specific disorders or medication affecting the central nervous system. In total, 32 Finnish children participated in the experiments. Of those three were excluded for the following reasons: two subjects due to excessive head movements and one subject due to low head position in the MEG helmet. The data included in the present study consisted of 29 children (mean age 8.17 years, SD: 1.05 years; 19 girls, 10 boys; 1 left-handed). All participants included had normal hearing as tested with an audiometry and normal or corrected-to-normal vision. This study was carried out in accordance with the recommendations of the Ethics Committee of the University of Jyväskylä. The protocol was approved by the Ethics Committee of the University of Jyväskylä. All children and their parents were informed about the project and they gave written consent in accordance with the Declaration of Helsinki to participate in the study. All subjects received gifts (movie tickets or shopping vouchers) as compensation for participation.

Stimuli and Task

The stimuli consisted of eight Finnish capital letters (A, E, I, O, U, Y, Ä, and Ö) written with Calibri font in black color and their corresponding speech sounds ([a], [e], [i], [o], [u], [y], [æ], and [ø]). Four categories of stimuli, auditory only (A), visual only (V), audiovisual congruent (AVC), and audiovisual incongruent (AVI) were presented in random order with 112 trials for each type of stimuli. The experiment was ca. 20 min in total with two short breaks. The duration of the auditory stimuli was 300 ms. The duration of the visual stimuli was 400 ms. For the audiovisual trials, the auditory and visual stimuli started at the same time. Each trial lasted 1500 ms and started with a fixation cross at the center of the screen for 500 ms, then followed by the presentation of auditory, visual or audiovisual stimuli (Figure 1). The visual stimuli were projected on the center of the screen in a gray background. The size of the visual stimuli was 0.6 cm × 0.6 cm for the fixation cross and 2 cm × 2 cm for the letters on a screen 1 m away from the participants. The sounds were delivered through insert earphones using MEG compatible lo-fi sound system at a comfortable loudness level. The stimuli were presented with Presentation (Neurobehavioral Systems, Inc., Albany, CA, United States) software running on a Windows computer. The experiment was conducted in a child-friendly environment in which we told a story of a cartoon character's adventure in a forest. In order to keep their attention equally on both auditory and visual stimuli, the participants were instructed to press a button using their right hand when they saw an animal drawing or heard an animal sound. In total eight animal drawings and their corresponding sounds were used as target stimuli and they occurred with 10% probability. Feedback (hit or miss) was given immediately after button press.



MEG and MRI

306-channel MEG data were recorded in a magnetically shielded room using Elekta Neuromag® TRIUX™ system (Elekta AB, Stockholm, Sweden) with 1000 Hz sampling rate and 0.1–330 Hz band-pass filter. The head position in relation to the sensors in the helmet was monitored continuously with five digitized head position indicator (HPI) coils attached to the scalp. Three HPI coils were placed on the forehead and one behind each ear. The position of HPI coils was determined in relation to three anatomic landmarks (nasion, left and right preauricular points) using the Polhemus Isotrak digital tracker system (Polhemus, Colchester, VT, United States) at the beginning of the recording. To allow the co-registration with individual magnetic resonance images (MRIs), an additional set of scalp points (>100) randomly distributed over the skull were also digitized. Electrooculogram (EOG) was recorded with two electrodes attached diagonally slightly below the left and slightly above the right eye and one ground electrode attached to the collarbone. The MEG was recorded in 68° upright gantry position.

Individual structural MR images were acquired from a private company offering MRI services (Synlab Jyväskylä). T1-weighted 3D-SE images were collected on a GE 1.5 T (GoldSeal Signa HDxt) MRI scanner using a standard head coil and with the following parameters: TR/TE = 540/10 ms, flip angle = 90°, matrix size = 256 × 256, slice thickness = 1.2 mm, sagittal orientation.

Behavioral Assessment

Cognitive skills were tested on a separate visit. The behavioral tests included the following: Wechsler Intelligence Scales for Children Third edition (Wechsler, 1991) and Wechsler Preschool and Primary Scales of Intelligence (Wechsler, 2003) for children above 6 years and for 6-year-olds, respectively. Block design (visuospatial reasoning), vocabulary (expressive vocabulary), and digit span (forward and backward; working memory) subtests were administered. In the block design test, the children are shown how to arrange blocks with red and white color to form a design and they have to build the same design. In more difficult sections the children are only shown the design in a figure and they have to build it. In the vocabulary test, the children hear a word and they have to describe the meaning of that word. In the digit span test, series of numbers are said to the participant and they have to repeat them either in forward or backward order. These tests were used to assess the children general cognitive skills and used as control variables for the possible associations between phonology and reading measures and the MEG indices.

Phonological awareness was tested using the Phonological processing task from NEPSY II (Korkman et al., 2007). In this task, the child is first asked to repeat a word and then to create a new word by leaving out a syllable or a phoneme, or by replacing one phoneme in the word with another phoneme.

Rapid automatized naming (Denckla and Rudel, 1976), in which pictures of five common objects or five letters had to be

named as quickly and as accurately as possible. The objects and letters were arranged in five rows each containing 15 objects. The task was audio-recorded and the time in seconds was calculated from the recording to be used in the analyses.

Three reading tests were included: word list reading using standardized test of word list reading (Häyrynen et al., 1999), number of correctly read words in 45 s was used as the score; non-word list reading based on Tests of Word Reading Efficiency (Torgesen et al., 1999), number of correctly read non-words in 45 s was used as the score; pseudoword text reading (Eklund et al., 2015), number of correctly read words and total reading time were used as the scores. Writing to dictation was also assessed in which the child heard 20 words and had to write them on a sheet of paper. Number of correctly written words was used as the score.

Data Analysis

Data were first processed with Maxfilter 3.0TM (Elekta AB) to remove external interference and correct for head movements. Bad channels were identified manually and were excluded and later reconstructed in Maxfilter. The temporal extension of the signal-space separation method (tSSS) was used in buffers of 30 s (Taulu et al., 2004; Taulu and Kajola, 2005; Taulu and Simola, 2006). Head position was estimated in 200 ms time windows and a 10 ms step for movement compensation. The MEG data were transformed to the mean head position across the recording session.

Data were then analyzed using MNE Python (0.15) (Gramfort et al., 2013). First, continuous MEG recordings were low-pass filtered at 40 Hz and epoched into -200 to 1000 ms trials relative to the stimulus onset. Data were then manually checked to remove any head movement-related artifacts and electronic jump artifacts. Then independent component analysis (ICA) using fastICA algorithm (Hyvärinen and Oja, 2000) was applied to remove eye blinks, horizontal eye movements and cardiac artifacts. MEG epochs exceeding 1 pT/cm for gradiometer or 3 pT for magnetometer peak-to-peak amplitudes were excluded from further analysis. Event-related fields were obtained by averaging trials in the four conditions separately. Sum of the auditory and visual response (A + V) was calculated by first equalizing the number of epochs between the unimodal conditions and then adding up the event-related fields of the auditory and visual only conditions. To match the noise level of A + V and AVC conditions and therefore to make these two conditions comparable, a subset of AVC trials was created by randomly selecting half the number of trials from the AVC condition which equates to the noise level in A + V condition.

Individual MRI were processed in Freesurfer (RRID: SCR_001847, Martinos Center for Biomedical Imaging, Charlestown, MA, United States) to obtain the cortical surface for source modeling. Three participants' MRIs were replaced by age and gender matched MRIs of other children (MRIs were not available for two children and the third one had a bad quality cortical surface reconstruction). Freesurfer reconstructed cortical surface was decimated to about 4098 evenly distributed vertices per hemisphere with 4.9 mm spacing. Cortically-constrained

and depth-weighted ($p = 0.8$) L2 minimum-norm estimate (wMNE) (Hämäläinen and Ilmoniemi, 1994; Lin et al., 2006) was calculated using one layer boundary element model (BEM) from the inner skull surface for all current dipoles with a loose orientation of 0.2. The noise covariance matrix was estimated from the raw 200 ms pre-stimulus baseline data over all conditions. For each current dipole, the estimated source amplitudes were calculated by taking the norm of the vectors. Source amplitudes were averaged within each label for the 68 brain regions defined by the Desikan-Killiany Atlas (Desikan et al., 2006). In order to capture the full extent of the sensory event-related field, the auditory source region was defined by a combination of superior temporal and transverse temporal brain areas and the visual source region was defined by a combination of lateral occipital, cuneus, pericalcarine and lingual brain areas. In addition, the fusiform area was defined as a region of interest for the N170 component based on previous studies (Cohen et al., 2000; Dehaene and Cohen, 2011).

In total, five auditory and visual event-related fields, the auditory N1m, N2m and late component, and the visual P1m and N170m were investigated in the present study. Peak latencies of these sensory responses were identified at sensor level (magnetometer) from the grand average of auditory and visual only conditions. The peak latencies were 109 ms (left) and 105 ms (right) for the auditory N1m, 241 ms (left) and 247 ms (right) for the auditory N2m and 448 ms (left) and 463 ms (right) for the auditory late component. The peak latencies were 104 ms (left) and 97 ms (right) for the visual P1m and 204 ms (left) and 192 ms (right) for the visual N170m. For all the four conditions (A, V, AVC, and AVI), the source level brain activities in the auditory or visual source regions were extracted by taking the average source activities of 50 ms time window centered around the peak latencies which were identified in the previous step. For auditory late component, a longer time window of 100 ms was used due to the extended time course of the response. In addition, individual peak latencies for each participant were also detected within the time window of each component in the source space.

Statistical Analysis

First, partial correlations (controlling for age in months) were calculated between the cognitive skill measures (see above) and the mean amplitudes and peak latencies of brain sensory responses in the four conditions using SPSS Statistics 24 software package (IBM Corp., Armonk, NY, United States). For the integration (A + V–AVC) and congruency (AVC–AVI) comparison, individual source waveforms in 68 brain regions extracted according to Desikan-Killiany atlas was used in nonparametric permutation (Maris and Oostenveld, 2007) *t*-tests with temporal clustering implemented in Mass Univariate ERP Toolbox (Groppe et al., 2011). The time window was selected from 0 to 1000 ms after stimulus onset and the number of permutations was 2000. The cluster alpha was 0.05 for both integration and congruency comparison. The family-wise *p* values were corrected for multiple comparisons. For regions that showed significant ($p < 0.05$) integration or congruency

effects, partial correlations (controlling for age in months) were calculated between cognitive skills and brain responses in each of these regions averaged in the time window of the significant clusters. In addition to the source amplitude values, a laterality index [(left-right)/(left+right)] was calculated for the activity from the fusiform gyrus to examine differences in the development of the hemispheric specialization to print as shown for example by (Maurer et al., 2008, 2010).

In addition, linear regression analyses were performed with cognitive skills as the dependent variable in SPSS Statistics 24. Children's age was entered first into the model followed by the brain responses as independent variables in order to determine if the different brain responses explain independent or overlapping portions of variance in the cognitive skills. Dependent and independent variables were selected based on significant partial correlations.

RESULTS

Cognitive Skills and Behavioral Performance

Descriptive statistics of the participants' cognitive skill measures and their behavioral performance in the cover task during MEG experiment are presented in **Table 1**.

Grand Averages

Grand averages of combined gradiometer channels in auditory only, visual only, audiovisual congruent and audiovisual incongruent conditions are shown in **Figure 2**. The waveforms were averaged over left and right temporal and occipital gradiometer channels (within the four circles shown in the sensor layout map).

The auditory and visual responses were identified in the magnetometer channels based on their topographies and timings. For the visualization purpose, the topography plot of auditory N1m, N2m and late component, and visual P1m and N170m are

shown at the local maximum of the global field power (GFP) in **Figure 3**.

Correlations Between Cognitive Skills and Sensory Brain Responses

No significant correlations were found between the scores in the cognitive skill measures for visuo-spatial reasoning (block design), general verbal skills (vocabulary), or verbal working memory (digit span) and the sensory brain responses. No significant correlations were found between age and the sensory brain responses.

Consistent correlations were found between the phonological processing accuracy, rapid naming speed of letters and auditory N1m, N2m, and LC responses (see **Tables 2–4**). No consistent correlation patterns were observed between peak latencies and cognitive skills (see **Supplementary Material**). In addition, the left auditory cortex activity at the late time window in response to the audiovisual stimuli showed rather systematic associations with phonology, rapid naming of letters and objects as well as non-word list reading accuracy. N170m amplitude in the left fusiform gyrus in the audiovisual conditions (both AVI and AVC) were significantly correlated with phonological processing. A similar correlation pattern was observed for the auditory only, audiovisual congruent and incongruent conditions in relation to cognitive skills thus indicating a high overlap between these brain measures. In general, we found that the larger the brain response the better the performance in the behavioral tasks for all of the correlations.

In the next step, linear regressions were used to predict the phonological and rapid naming (the dependent variable) using age and the brain responses that showed significant partial correlations as predictors (independent variables). Age was entered first into the model followed first by the significant auditory variables and visual variables using stepwise method and finally by the significant audiovisual variables also using the stepwise method. This model was used to disentangle possible overlapping variance explained by auditory/visual and audiovisual brain responses. In the multiple regression model, as shown in **Table 5** the auditory late component from the left hemisphere was the only significant predictor of the phonological skills and RAN letters.

The scatterplots (**Figure 4**) show that, in general, the larger the source activity in the auditory cortex the more likely it is that the child has better phonological processing skills and faster rapid naming abilities.

Integration and Congruency Effects Integration Effect (A + V vs. AVC)

Cluster-based permutation tests showed that audiovisual integration effect was found in multiple brain regions in the parietal and temporal areas after ca. 250 ms ($p < 0.05$) as shown in **Figure 5**. In total eight significant clusters were found in eight brain regions of the Desikan-Killiany atlas. These clusters were in the left (317–499 ms) and right (315–818 ms) inferior parietal, left (391–585 ms) and right (306–797 ms) supramarginal, right (271–529 ms) precuneus, right (551–755 ms) postcentral and

TABLE 1 | Descriptive statistics of the participants' cognitive skill measures and behavioral performance in the cover task during MEG experiment ($N = 29$).

	Mean	SD	Range
Age (years)	8.17	1.05	6–11
Block design, raw score	25.53	9.50	10–52
Vocabulary, raw score	28.17	8.71	12–50
Digit span, raw score	11.67	1.88	8–16
NEPSY phonological processing, raw score	42.53	5.74	30–52
Rapid automatic naming(letters), time (s)	40.01	11.20	26–69
Rapid automatic naming(objects), time (s)	60.61	12.63	38–90
Word list reading, number of correct items	58.46	26.32	10–104
Non-word list reading, number of correct items	31.39	14.23	7–67
Non-word text reading, time (s)	118.32	74.74	36–390
Non-word text reading, number of correct words	29.96	5.33	16–37
Writing to dictation, number of correct words	34.07	8.14	10–40
MEG cover task, accuracy, %	96.91	3.85	81–100
MEG cover task, reaction time, ms	642	127	475–1064

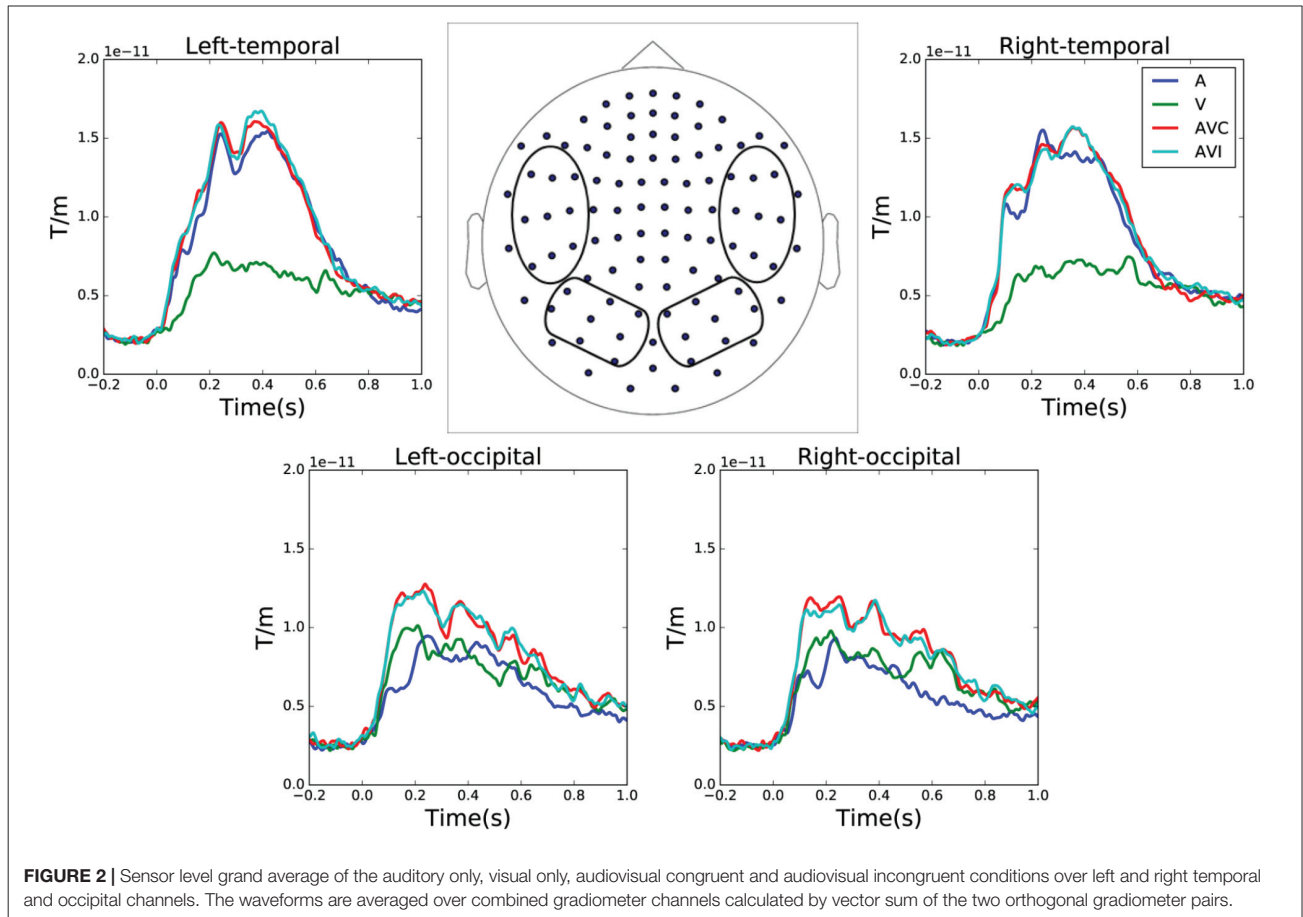


FIGURE 2 | Sensor level grand average of the auditory only, visual only, audiovisual congruent and audiovisual incongruent conditions over left and right temporal and occipital channels. The waveforms are averaged over combined gradiometer channels calculated by vector sum of the two orthogonal gradiometer pairs.

right superior (535–827 ms), and middle (346–749 ms) temporal cortices.

Congruency Effect (AVC vs. AVI)

Cluster-based permutation test did not reveal significant effects ($p > 0.05$) in congruency comparison.

Correlations Between Cognitive Skills and the Brain Activity Related to Multimodal Integration

The difference between the AVC and A + V conditions was calculated and the average source amplitudes from the different brain regions in the time window identified by the permutation test were used for the correlation analyses with cognitive skills (Table 6). Representative partial correlations between suppressive integration and behavioral tests are shown in Figure 6.

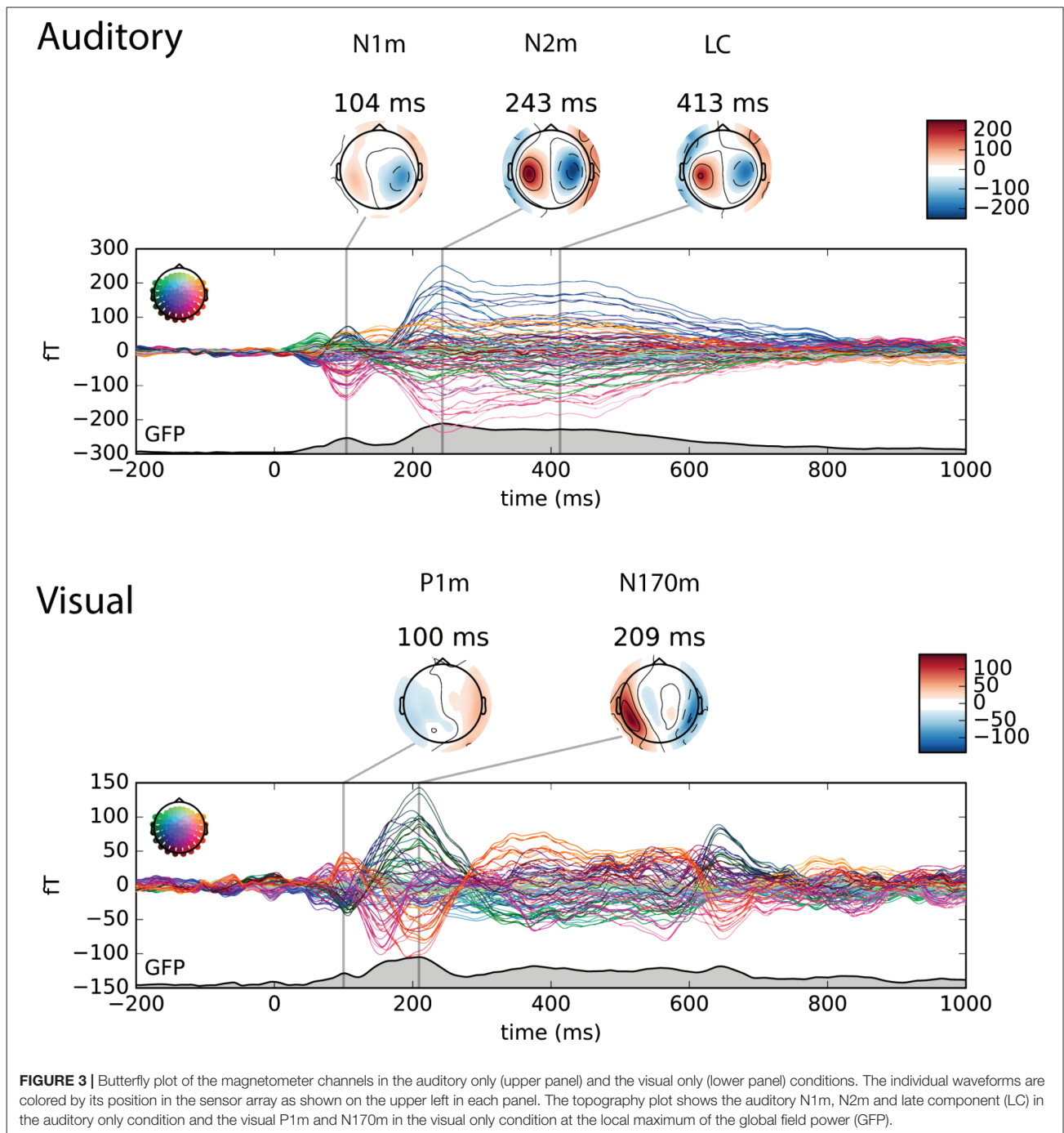
DISCUSSION

In this study, auditory and visual responses, as well as audiovisual integration of letters and speech sounds were correlated with children’s behavioral cognitive skills. Results from the current study revealed that auditory processing, especially the auditory

processing in the late time window was the driving force for the correlation between sensory evoked fields and phonological skills. The visual N170 in the left fusiform gyrus in the audiovisual condition was also correlated with phonological skills. In addition, audiovisual suppressive integration was localized mainly in the temporoparietal brain regions and showed an independent contribution from the sensory evoked fields to the reading skills.

It has been shown that the sequence of activation in response to speech sounds is strikingly different in children compared with adults (Wunderlich et al., 2006; Parviainen et al., 2011). Children showed prolonged responses to sound with a major peak at 250 ms in both left and right hemispheres (Parviainen et al., 2011) while a corresponding effect occurred about 100 ms specifically in the left hemisphere in adults (Parviainen et al., 2005). This matches with the current findings that showed a major negative going peak around 250 ms after speech sound onset. The response at 250 ms is usually followed by a second activity peak around 400 ms (Ceponiene et al., 2001, 2005, 2008).

The auditory late component seems to be sensitive to the speech sounds as can be seen from the study in children in which a strong late activation around 400 ms was observed in speech sounds compared to other types of sounds (Parviainen et al., 2011). The activity during the late component time



window has been suggested in other studies to be related to late stages of phonological processing (Stevens et al., 2013; Bann and Herdman, 2016) or to orthographic-phonological mapping (Weber-Fox et al., 2003). However, in our study the late processing (around 413 ms) seemed to be linked to the auditory stimuli. This fits with previous studies suggesting that this time window could reflect the effect of speech sound representations (Szymanski et al., 1999; Ceponiene et al., 2001, 2005; Kuuluvainen

et al., 2016) and it is sensitive to phonological priming (Bonte and Blomert, 2004). The response has also been suggested to be important for receptive language processing (Ceponiene et al., 2008) which also matched with the correlation pattern of the current study. Overall this could imply that the later stages of integrative speech sound processing are important also for learning to read and for phonological skills. Although the activity around 400 ms seems to mature early in development

TABLE 2 | Partial correlations (controlling for age) between cognitive skills and the auditory responses in the auditory only condition.

Auditory	Phonological processing	RAN letters	RAN objects	Word list reading	Non-word list reading	Non-word text time	Non-word text accuracy	Writing to dictation
L, N1m	0.363	-0.219	-0.173	-0.017	0.078	-0.037	0.143	0.141
R, N1m	0.384*	-0.312	-0.368	0.304	0.375	-0.226	0.305	0.286
L, N2m	0.454*	-0.349	-0.189	0.203	0.323	-0.216	0.312	0.142
R, N2m	0.286	-0.228	-0.265	0.279	0.316	-0.204	0.301	0.273
L, LC	0.499**	-0.399*	-0.358	0.215	0.308	-0.181	0.241	0.234
R, LC	0.472*	-0.338	-0.280	0.187	0.258	-0.164	0.284	0.313

Note: * $p < 0.05$, ** $p < 0.01$.

The auditory components are left (L) and right (R) auditory N1m, N2m and late component (LC).

TABLE 3 | Partial correlations (controlling for age) between cognitive skills and the visual responses in the visual only condition.

Visual	Phonological processing	RAN letters	RAN objects	Word list reading	Non-word list reading	Non-word text time	Non-word text accuracy	Writing to dictation
L, VC, P1	0.343	-0.171	-0.005	-0.241	-0.206	0.255	-0.053	-0.015
R, VC, P1	0.372	-0.212	-0.116	-0.136	-0.081	0.195	0.093	0.075
L, FG, N170	0.282	-0.131	0.036	-0.078	0.011	0.023	0.113	0.027
R, FG, N170	0.218	-0.152	-0.150	-0.080	-0.029	0.079	0.094	0.074
V, LI	0.074	0.062	0.211	-0.017	0.004	-0.125	-0.103	-0.058

Note: VC, visual cortices; FG, fusiform gyrus; LI, laterality index.

The visual responses are left (L) and right (R) visual P1m in the visual cortices and N170m in the fusiform gyrus.

TABLE 4 | Partial correlations (controlling for age) between cognitive skills and the auditory and visual responses in the audiovisual conditions [the first row in each cell is audiovisual congruent (AVC) and the second row audiovisual incongruent (AVI)].

AVC AVI	Phonological processing	RAN letters	RAN objects	Word list reading	Non-word list reading	Non-word text time	Non-word text accuracy	Writing to dictation
L, AC, N1	0.272	-0.162	-0.124	-0.074	-0.047	0.091	0.098	0.154
	0.422*	-0.250	-0.215	-0.025	0.027	-0.014	0.106	0.182
R, AC, N1	0.351	-0.266	-0.315	0.299	0.390*	-0.213	0.320	0.257
	0.337	-0.253	-0.352	0.267	0.337	-0.180	0.314	0.253
L, AC, N2	0.420*	-0.311	-0.265	0.245	0.378	-0.277	0.229	0.119
	0.338	-0.290	-0.250	0.284	0.346	-0.280	0.235	0.054
R, AC, N2	0.278	-0.247	-0.266	0.290	0.329	-0.230	0.262	0.252
	0.235	-0.203	-0.241	0.216	0.305	-0.198	0.272	0.259
L, AC, LC	0.506**	-0.412*	-0.405*	0.311	0.395*	-0.288	0.252	0.272
	0.441*	-0.381*	-0.394*	0.218	0.293	-0.251	0.153	0.227
R, AC, LC	0.448*	-0.312	-0.235	0.107	0.126	-0.122	0.193	0.267
	0.456*	-0.309	-0.241	0.151	0.201	-0.158	0.255	0.293
L, VC, P1	0.292	-0.135	0.010	-0.198	-0.149	0.188	-0.062	0.004
	0.243	-0.087	0.067	-0.308	-0.236	0.281	-0.084	-0.110
R, VC, P1	0.315	-0.197	-0.145	0.015	0.066	0.001	0.057	0.127
	0.297	-0.210	-0.145	-0.024	0.015	0.031	0.091	0.112
L, FG, N170	0.404*	-0.261	-0.090	0.014	0.124	-0.011	0.202	0.088
	0.427*	-0.280	-0.113	0.064	0.203	-0.025	0.255	0.122
R, FG, N170	0.213	-0.133	-0.173	0.020	0.072	-0.046	0.121	0.191
	0.221	-0.165	-0.235	0.032	0.080	-0.036	0.113	0.170
AV, LI	0.208	-0.079	0.169	-0.050	-0.052	0.045	-0.065	-0.160
	0.220	-0.093	0.239	-0.058	-0.034	0.043	0.010	-0.076

Note: AC, auditory cortices; VC, visual cortices; FG, fusiform gyrus; LI, laterality index. * $p < 0.05$, ** $p < 0.01$.

The auditory components are left (L) and right (R) auditory N1m, N2m and late components. The visual components are left (L) and right (R) visual P1m in the visual cortices and N170m in the fusiform gyrus.

TABLE 5 | Linear regression analysis using phonological and rapid naming as the dependent variable, age was entered first in the model, then the brain responses that showed significant partial correlations as predictors (independent variables).

Cognitive skills	Step	Standardized beta	Δ R2
Phonological processing	1. age	0.256	0.091
	2. Left auditory LC	0.487**	0.235**
	3. AVC/AVI responses	ns	ns
RAN letters	1. age	-0.059	0.009
	2. Left auditory LC	-0.405*	0.163*
	3. AVC/AVI responses	ns	ns

Note: Beta = standardized Beta coefficient; R2 change = unique variance accounted for at each step of the 2-step (enter method for age; stepwise for brain measures) multiple regression analyses. *p < 0.05, **p < 0.01.

(Kushnerenko et al., 2002), there is still substantial variation in the response amplitude at school-age that is systematically linked with cognitive skills related to language processing.

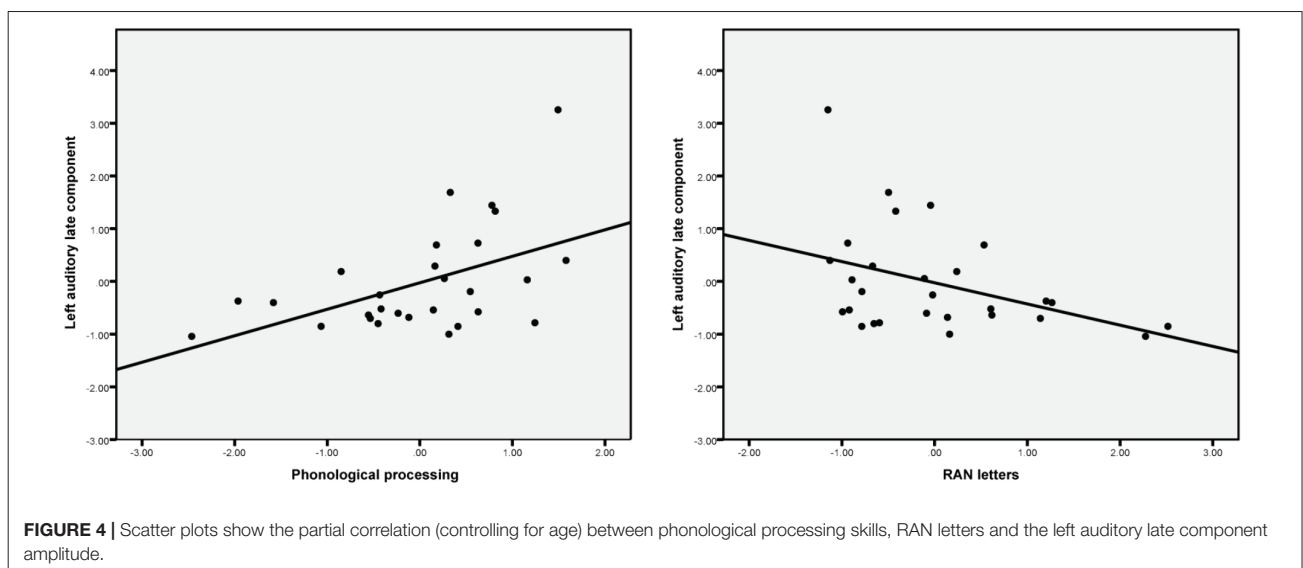
In the current study, we found correlations between N1, N2, late component and phonological processing for both auditory and AV conditions. Although the regression analysis showed that only the left auditory late component explains unique variance among the brain measures implicating that the early responses do not have independent variance from the late activity that is related to phonological processing. From the time window around 100 ms correlations have been found between brain responses and preschool cognitive skills also in other studies. For example, auditory P1 response has been shown in typically developed children to be associated with phonological and pre-reading skills (Kuuluvainen et al., 2016). In addition, for children at risk for dyslexia, their P1 response amplitudes elicited by speech sound stimuli were smaller compared to controls (Lovio et al., 2010). Similarly, in a study (Hämäläinen et al., 2015) investigating the event-related potentials to tones in children with multiple risk

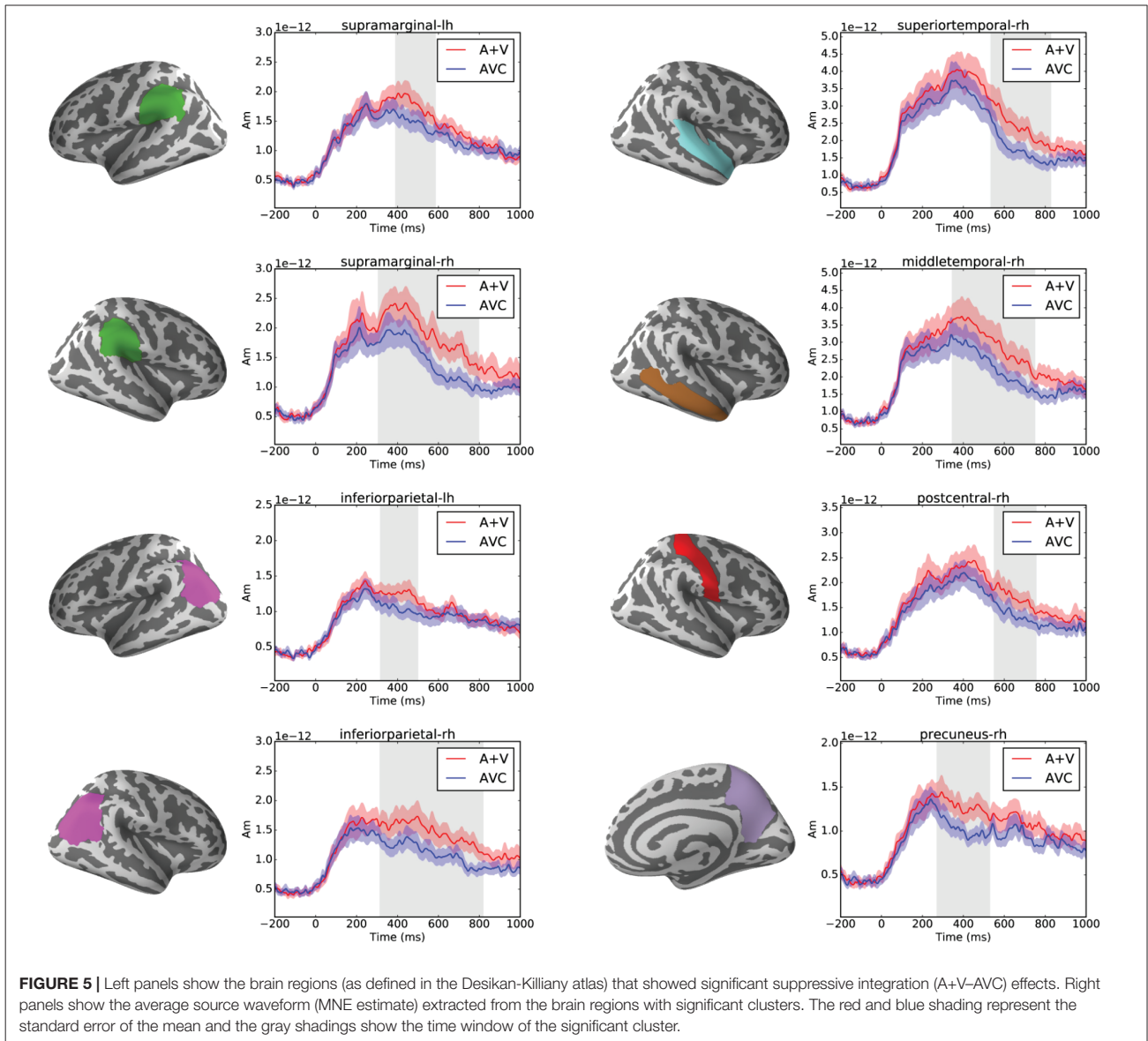
factors for dyslexia, the amplitudes at the P1–N2 time window was correlated with letter knowledge and phonological skills. The N1 and N2 time window has also been shown to be sensitive to reading level differences in response to phonological priming (Bonte, 2004) and nonspeech stimuli (Espy et al., 2004).

The N2 response has been linked to reading and reading-related skills in previous studies. For example, the N2m has been found to correlate with reading skills in children (Parviainen et al., 2011) and the N2 response has been reported to have larger amplitudes in response to speech and non-speech sounds in dyslexic children compared with control group and such enhanced brain responses were correlated with reading skills (Hämäläinen et al., 2013). Furthermore, the brain activity at the N2 time window has been found to correlate with phonological skills, as well as reading and writing accuracy in children with dyslexia (Lohvansuu et al., 2014). The N2m response strength in the left hemisphere in the current study was correlated with phonological skills further supporting the hypothesis that this time window is important to language-related skill development.

We also found a significant correlation between rapid naming ability and auditory late component amplitude. Previous research (Kuuluvainen et al., 2016) showed a similar relationship between N4 and rapid naming speed in preschool children in which N4 was suggested to be linked to accessing phonological representations. Overall, the correlation patterns found in the current study between the phonological and rapid naming ability and auditory brain responses are consistent with and in support of the earlier literature.

Audiovisual responses shared a large portion of variance with the auditory responses, and furthermore, both showed an association with phonology. In order to disentangle contributions of the auditory processing from the audiovisual processing, we run regression analyses with both auditory and audiovisual brain responses as predictors. No unique variance was left to





be explained by the responses to the audiovisual stimuli on the phonological skills after the left auditory late response was taken into account. The regression analyses thus showed the auditory response to be the driving force behind the association with phonological skills.

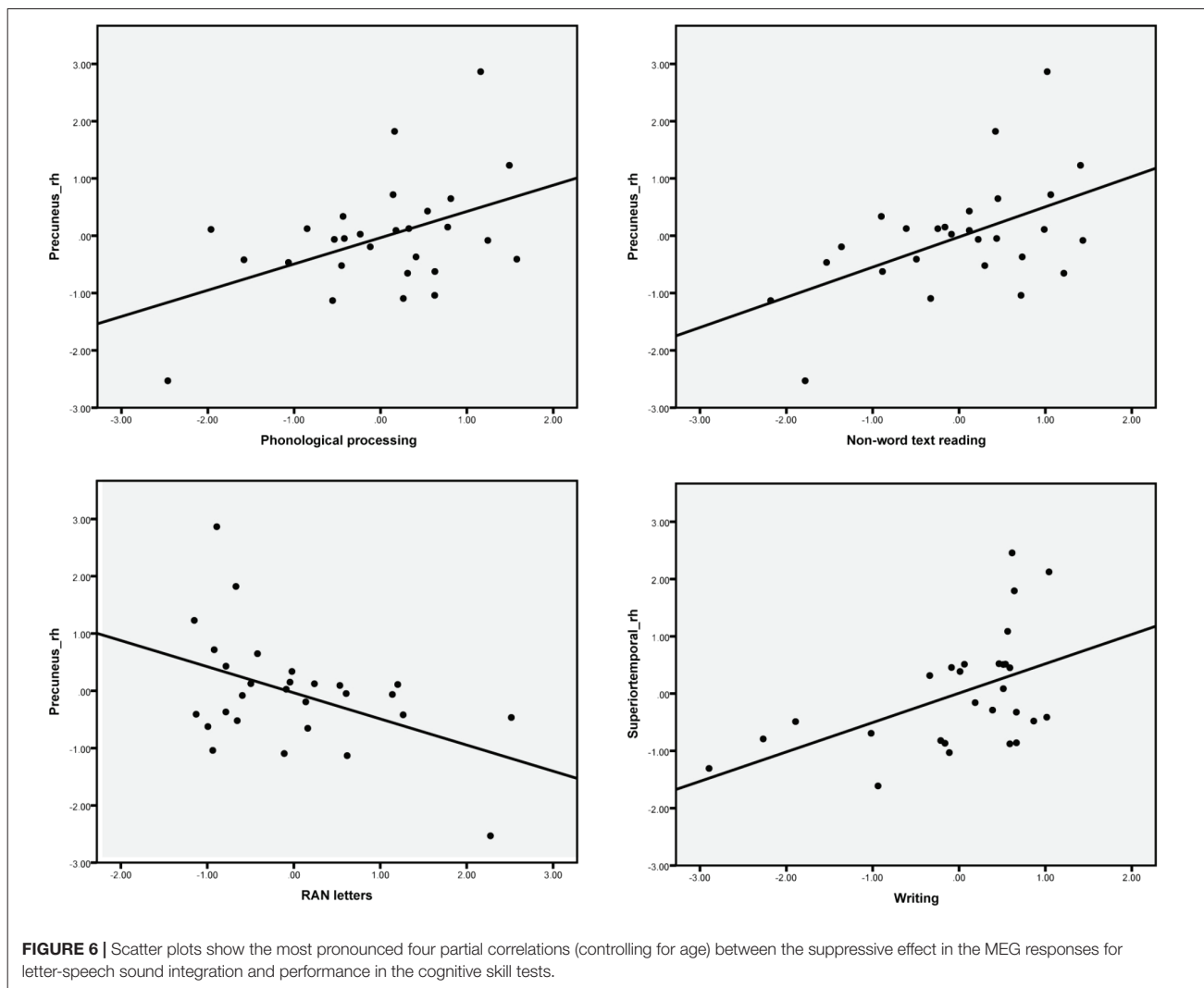
N170 amplitude and the laterality index of the N170 were not significantly correlated with any of the cognitive skills in the visual only condition. Most previous studies (Cohen et al., 2000; Dehaene et al., 2002; Maurer et al., 2005, 2008; Dehaene and Cohen, 2011) found brain specialization for letter strings and whole words in VWFA (as indexed by N170 responses in EEG/MEG). Presentation of single letters in our study instead of letter strings or words could therefore have led to the lack of findings for the N170 response in the visual only condition. However, previous studies (McCandliss

and Noble, 2003; Maurer et al., 2010) have suggested the left lateralization of N170 for words to be partly driven by an automatic link between orthographic and phonological systems. Interestingly, the N170 response showed significant correlation with phonological skills in both audiovisual congruent and incongruent conditions in the left fusiform area. This result could suggest a possible top-down feedback activation of the VWFA and the lateral inferior temporal cortex from auditory and audiovisual integration sites. It has been reported that the VWFA could be activated during speech processing through a top-down modulation (Dehaene et al., 2010; Desroches et al., 2010; Yoncheva et al., 2010). Such auditory/audiovisual processing modulation fits well with the significant correlation between phonological processing and N170 responses in left fusiform in the audiovisual conditions in our study. Similar results were

TABLE 6 | Partial correlations (controlling for age) between cognitive skills and averages of the brain responses in the regions and time windows where significant audiovisual integration effects were revealed by the cluster-based permutation analyses.

A+V-AVC	Phonological processing	RAN letters	RAN objects	Word list reading	Non-word list reading	Non-word text time	Non-word text accuracy	Writing to dictation
Supramarginal_lh	0.365	-0.421*	-0.167	0.163	0.233	-0.130	0.297	0.006
Supramarginal_rh	0.324	-0.262	-0.067	0.099	0.096	0.079	0.300	0.293
Inferiorparietal_lh	-0.055	-0.056	0.097	-0.125	-0.036	-0.109	-0.166	-0.136
Inferiorparietal_rh	0.401*	-0.353	-0.263	0.192	0.269	-0.126	0.266	0.328
Superiortemporal_rh	0.229	-0.196	-0.259	0.277	0.304	-0.135	0.211	0.504**
Middletemporal_rh	0.304	-0.215	-0.258	0.306	0.371	-0.230	0.210	0.408*
Postcentral_rh	0.215	-0.131	0.104	-0.007	0.133	0.031	0.262	0.263
Precuneus_rh	0.457*	-0.455*	-0.317	0.435*	0.405*	-0.186	0.519**	0.376

Note: * $p < 0.05$, ** $p < 0.01$.



found in an MEG study in which occipitotemporal letter-string-sensitive activation strength was also reported to be correlated with phonological skills in children (Parviainen et al., 2006).

When comparing the summed unimodal responses to the audiovisual responses, suppressive audiovisual integration effect was found in right temporal and both left and right parietal regions. These regions partly match with a previous MEG

study (Raij et al., 2000) in adults about LSS integration in which a suppressive integration effect was found in the right temporo-occipito-parietal junction and the left and right STS. In the current study, we found suppressive audiovisual integration effects mostly in the temporoparietal areas but not in the frontal cortices reported in (Raij et al., 2000). This could be due to the difference in the experimental design since an active implicit audiovisual task was used in our study whereas (Raij et al., 2000) used an active explicit matching task, which could recruit more top-down task related audiovisual attention processes (van Atteveldt et al., 2007). The dorsal (temporoparietal) system, including supramarginal gyrus/angular gyrus in the inferior parietal lobule and the posterior superior temporal gyrus (pSTG) is thought to be related to mapping visual print onto the phonological and semantic structures of language (Sandak et al., 2004). Compared with the rather consistent findings in the superior temporal cortex for LSS integration in adults (Raij et al., 2000; van Atteveldt et al., 2004), it seems that the early readers have recruited more widely distributed temporoparietal cortical networks to support learning the association of orthography with phonological codes (Pugh et al., 2013). The suppressive LSS integration effect in the parietal areas at the rather late time window could be related to top-down modulation of the audiovisual processing and reflect less automatic processing of the stimuli than in adults. Pugh et al. (2013) also find a similar correlation between BOLD response and reading skills in the precuneus, which is similar to the current study, and they interpret their finding as part of the visual attention network that seems to impact reading development. They also suggest that this could reflect the integration between visual, language and attentional processes. Lack of the suppressive integration effect at the left superior temporal areas could be related to the less automatic processing of the multimodal stimuli in early readers (Froyen et al., 2009; Blomert, 2011).

The timing of this integration effect was mostly from about 300 to 600 ms in the present study, which matches well with the previous studies using similar stimuli and paradigms (Raij et al., 2000; Herdman et al., 2006; Jost et al., 2014). The relatively late time window is probably due to the fact that bimodal audiovisual integration happens after the early unimodal processing of sound in the auditory cortex and print in the visual cortex (Raij et al., 2000) and possibly involve the feedback projection to auditory cortex in a late stage of processing (van Atteveldt et al., 2004).

Significant partial correlations were found between the audiovisual integration effect and phonological skills, rapid naming abilities as well as reading and writing skills. Phonological skills were correlated with the strength of the audiovisual integration effect in the right inferior parietal and precuneus regions, while rapid naming of letters was correlated with the strength of the audiovisual integration in the left supramarginal and right precuneus regions. Previously research has found similar associations between both structural (gray matter volume indices) (Raschle et al., 2011) and functional (Raschle et al., 2012) changes in these temporoparietal regions and pre-reading skills such as phonology and rapid naming. Moreover, activations in

left parietal (angular gyrus) lobe was correlated with individual at-risk index scores for dyslexia in pre-readers (Specht et al., 2009). Reduced LSS is suggested to be linked to a deficit in auditory processing of speech sounds, which in turn predicts phonological skills (Blau et al., 2009). Consistent correlation was found between the strength of the audiovisual integration effect in the right precuneus and reading skills such as word list, nonword list and nonword text reading accuracy. This matches well with results from one recent study which used similar brain-behavior correlation analysis with fMRI and showed the activation in the precuneus to print and speech sounds of words and pseudowords to be correlated with reading-related skills (Pugh et al., 2013). Finally, writing skills were also significantly correlated with the strength of the audiovisual integration effect in the right superior and middle temporal regions. This might suggest that the skills required in writing to dictation are more associated with auditory processes for speech than those required for reading (Hämäläinen et al., 2009). Taken together, these results highlight the important role of LSS in the temporoparietal area in early reading acquisition (Blomert and Froyen, 2010; Blomert and Willems, 2010).

Audio-visual congruency did not produce significant effects in the brain responses in the present study. Here we discuss possible reasons for this. First, the congruency effect which heavily depends on the task demands (Andersen et al., 2004; van Atteveldt et al., 2007), also seems to interact with the brain imaging method (fMRI vs. MEG). For example, several previous fMRI studies on children have found a congruency effect using similar implicit active tasks to ours (Blau et al., 2010; Brem et al., 2010). In contrast, use of an active explicit matching task in fMRI has been reported to overrule the congruency effect (van Atteveldt et al., 2007). However, the MEG study of (Raij et al., 2000) used an active task forcing the participants to relate letters to sounds and reported an audiovisual congruency effect in the heteromodal superior temporal cortex. Therefore, it seems that the task demands modulate differently the MEG and BOLD responses. Second, it is also possible that the children in the present study may not establish fully automatized LSS integration as many of them only have 1 or 2 years of reading instruction. Previous research (Froyen et al., 2009; Blomert, 2011) using MMN paradigm has shown the protracted developmental trajectory of LSS integration and this may be reflected in the absence of congruency effect in the present study. Finally, almost all previous electrophysiological studies (Froyen et al., 2008, 2010; Žarić et al., 2014) examining letter-speech sound congruency in children have used an oddball paradigm, it is likely that congruency is pronounced in the oddball paradigm, but not in the simple LSS paradigm used in the present study. The audiovisual integration and congruency comparisons indicated that children seemed to utilize more general multimodal integration processes of letters and speech sounds, but have not reached the fully automatic level of integration as shown by the absent of congruency effect.

A cohort of beginning readers with relatively wide age range (6–11 years) was recruited to examine the reading and reading related cognitive skills as continuums. Even though we controlled for age in all of the correlation and regression analyses, age did

not seem to have a large impact on the results. This finding is similar to that of, for example, the study by Pugh et al. (2013). It seems that the correlations were driven more by learning of these cognitive skills than general maturation of the central nervous system.

According to the general neurodevelopmental theory for reading proposed by (Pugh et al., 2000a; Cornelissen et al., 2010), the temporal and dorsal parietal networks are crucial for the early stage of reading acquisition. Working together with the anterior regions (especially the inferior frontal gyrus), the dorsal (temporoparietal) reading system is involved in the emergence of phonological awareness (Katzir et al., 2005) and in forming associations between orthography, phonology, and semantics (Pugh et al., 2001). Such associations will then shape the organization and connectivity of left occipitotemporal regions including the VWFA (Dehaene and Cohen, 2011) for supporting fluent reading in advanced readers. The present study highlighted the important role of the temporoparietal route in developing phonological awareness and forming automatic LSS in early readers.

A possible concern regarding our study relates to the accuracy of MEG source reconstruction in children, which could be affected by many factors including the relatively large distance of the child's head to the MEG sensors, imprecise cortical surface reconstruction, suboptimal forward and inverse solution parameters for the child brain and potential MEG-MRI coregistration errors. These could lead to misallocation of brain activity to neighboring brain regions from their true locations in the source analyses. In general, we followed the recommended analysis practice proposed by (Jas et al., 2017) and checked in each step the quality of the data carefully. Furthermore, MEG is less sensitive to the conductivity parameters of the head tissues than EEG which should allow better reconstruction of source activity in children. In addition, we used relatively large brain regions, and in the case of LSS integration effects whole brain analysis, capturing most of the brain activity in the different conditions taking into account possible limitations in localization accuracy of the brain activity.

CONCLUSION

In conclusion, brain-behavior analyses were used to explore the relationship between behavioral tasks measuring different

cognitive skills and brain responses related to auditory and visual processing of letters and speech sounds in beginning readers. Regression analysis identified the auditory late component in response to speech sounds to be the most significant predictor of phonological skills and rapid naming. In addition, the audiovisual integration effect was found in left and right temporoparietal regions and several of these temporal and parietal regions showed contribution to reading and writing skills. Findings from the current study point to the important role of temporoparietal regions in learning letter-speech sound associations in early reading development. A more detailed neurocognitive model, including additional measures such as functional connectivity, is needed for better understanding of the cortical organization and the developmental trajectory of LSS in children learning to read.

AUTHOR CONTRIBUTIONS

WX, JH, and OK designed the study. WX, JH, and OK performed the MEG experiments. WX, JH, and SM analyzed the data. All authors discussed the results and contributed to the final manuscript.

FUNDING

This work was supported by the European Union projects ChildBrain (Marie Curie Innovative Training Networks, no. 641652), Predictable (Marie Curie Innovative Training Networks, no. 641858), and the Academy of Finland (MultiLeTe #292466).

ACKNOWLEDGMENTS

We would like to thank Katja Koskialho, Ainomaija Laitinen, and Sonja Tiri for their help in data collection.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fnhum.2018.00304/full#supplementary-material>

REFERENCES

- Albrecht, R., Suchodoletz, W., and Uwer, R. (2000). The development of auditory evoked dipole source activity from childhood to adulthood. *Clin. Neurophysiol.* 111, 2268–2276. doi: 10.1016/S1388-2457(00)00464-8
- Andersen, T. S., Tiippana, K., and Sams, M. (2004). Factors influencing audiovisual fission and fusion illusions. *Brain Res. Cogn. Brain Res.* 21, 301–308. doi: 10.1016/j.cogbrainres.2004.06.004
- Bach, S., Richardson, U., Brandeis, D., Martin, E., and Brem, S. (2013). Print-specific multimodal brain activation in kindergarten improves prediction of reading skills in second grade. *Neuroimage* 82, 605–615. doi: 10.1016/j.neuroimage.2013.05.062
- Bann, S. A., and Herdman, A. T. (2016). Event related potentials reveal early phonological and orthographic processing of single letters in letter-detection and letter-rhyme paradigms. *Front. Hum. Neurosci.* 10:176. doi: 10.3389/fnhum.2016.00176
- Besle, J., Fischer, C., Bidet-Caulet, A., Lecaigard, F., Bertrand, O., and Giard, M.-H. (2008). Visual activation and audiovisual interactions in the auditory cortex during speech perception: intracranial recordings in humans. *J. Neurosci.* 28, 14301–14310. doi: 10.1523/JNEUROSCI.2875-08.2008
- Besle, J., Fort, A., Delpuech, C., and Giard, M.-H. (2004). Bimodal speech: early suppressive visual effects in human auditory cortex. *Eur. J. Neurosci.* 20, 2225–2234. doi: 10.1111/j.1460-9568.2004.03670.x
- Binder, J. R., Desai, R. H., Graves, W. W., and Conant, L. L. (2009). Where is the semantic system? A critical review and meta-analysis of 120 functional

- neuroimaging studies. *Cereb. Cortex* 19, 2767–2796. doi: 10.1093/cercor/bhp055
- Blau, V., Reithler, J., Van Atteveldt, N., Seitz, J., Gerretsen, P., Goebel, R., et al. (2010). Deviant processing of letters and speech sounds as proximate cause of reading failure: a functional magnetic resonance imaging study of dyslexic children. *Brain* 133, 868–879. doi: 10.1093/brain/awp308
- Blau, V., van Atteveldt, N., Ekkebus, M., Goebel, R., and Blomert, L. (2009). Reduced neural integration of letters and speech sounds links phonological and reading deficits in adult dyslexia. *Curr. Biol.* 19, 503–508. doi: 10.1016/j.cub.2009.01.065
- Blau, V., van Atteveldt, N., Formisano, E., Goebel, R., and Blomert, L. (2008). Task-irrelevant visual letters interact with the processing of speech sounds in heteromodal and unimodal cortex. *Eur. J. Neurosci.* 28, 500–509. doi: 10.1111/j.1460-9568.2008.06350.x
- Blomert, L. (2011). The neural signature of orthographic-phonological binding in successful and failing reading development. *Neuroimage* 57, 695–703. doi: 10.1016/j.neuroimage.2010.11.003
- Blomert, L., and Froyen, D. (2010). Multi-sensory learning and learning to read. *Int. J. Psychophysiol.* 77, 195–204. doi: 10.1016/j.ijpsycho.2010.06.025
- Blomert, L., and Willems, G. (2010). Is there a causal link from a phonological awareness deficit to reading failure in children at familial risk for dyslexia? *Dyslexia* 16, 300–317. doi: 10.1002/dys.405
- Bonte, M. (2004). Developmental changes in ERP correlates of spoken word recognition during early school years: a phonological priming study. *Clin. Neurophysiol.* 115, 409–423. doi: 10.1016/s1388-2457(03)00361-4
- Bonte, M. L., and Blomert, L. (2004). Developmental dyslexia: ERP correlates of anomalous phonological processing during spoken word recognition. *Brain Res. Cogn. Brain Res.* 21, 360–376. doi: 10.1016/j.cogbrainres.2004.06.010
- Booth, J. R., Burman, D. D., Meyer, J. R., Gitelman, D. R., Parrish, T. B., and Mesulam, M. M. (2003). Relation between brain activation and lexical performance. *Hum. Brain Mapp.* 19, 155–169. doi: 10.1002/hbm.10111
- Brandwein, A. B., Foxe, J. J., Russo, N. N., Altschuler, T. S., Gomes, H., and Molholm, S. (2011). The development of audiovisual multisensory integration across childhood and early adolescence: a high-density electrical mapping study. *Cereb. Cortex* 21, 1042–1055. doi: 10.1093/cercor/bhq170
- Brem, S., Bach, S., Kucian, K., Guttorm, T. K., Martin, E., Lyytinen, H., et al. (2010). Brain sensitivity to print emerges when children learn letter-speech sound correspondences. *Proc. Natl. Acad. Sci. U.S.A.* 107, 7939–7944. doi: 10.1073/pnas.0904402107
- Buchsbam, B. R., and D'Esposito, M. (2008). The search for the phonological store: from loop to convolution. *J. Cogn. Neurosci.* 20, 762–778. doi: 10.1162/jocn.2008.20501
- Calvert, G. A., and Thesen, T. (2004). Multisensory integration: methodological approaches and emerging principles in the human brain. *J. Physiol. Paris* 98, 191–205. doi: 10.1016/j.jphysparis.2004.03.018
- Ceponiene, R., Alku, P., Westerfield, M., Torki, M., and Townsend, J. (2005). ERPs differentiate syllable and nonphonetic sound processing in children and adults. *Psychophysiology* 42, 391–406. doi: 10.1111/j.1469-8986.2005.00305.x
- Ceponiene, R., Shestakova, A., Balan, P., Alku, P., Yiaguchi, K., and Naatanen, R. (2001). Children's auditory event-related potentials index sound complexity and "speechness." *Int. J. Neurosci.* 109, 245–260. doi: 10.3109/00207450108986536
- Ceponiene, R., Torki, M., Alku, P., Koyama, A., and Townsend, J. (2008). Event-related potentials reflect spectral differences in speech and non-speech stimuli in children and adults. *Clin. Neurophysiol.* 119, 1560–1577. doi: 10.1016/j.clinph.2008.03.005
- Cohen, L., Dehaene, S., Naccache, L., Lehéry, S., Dehaene-Lambertz, G., Hénaff, M. A., et al. (2000). The visual word form area: spatial and temporal characterization of an initial stage of reading in normal subjects and posterior split-brain patients. *Brain* 123(Pt 2), 291–307. doi: 10.1093/brain/123.2.291
- Cornelissen, P., Hansen, P., Kringelbach, M., and Pugh, K. (2010). *The Neural Basis of Reading*. Oxford: Oxford University Press. doi: 10.1093/acprof:oso/9780195300369.001.0001
- Dehaene, S., and Cohen, L. (2011). The unique role of the visual word form area in reading. *Trends Cogn. Sci.* 15, 254–262. doi: 10.1016/j.tics.2011.04.003
- Dehaene, S., Cohen, L., Morais, J., and Kolinsky, R. (2015). Illiterate to literate: behavioural and cerebral changes induced by reading acquisition. *Nat. Rev. Neurosci.* 16, 234–244. doi: 10.1038/nrn3924
- Dehaene, S., Le Clec'h, G., Poline, J.-B., Le Bihan, D., and Cohen, L. (2002). The visual word form area: a prelexical representation of visual words in the fusiform gyrus. *Neuroreport* 13, 321–325. doi: 10.1097/00001756-200203040-00015
- Dehaene, S., Pegado, F., Braga, L. W., Ventura, P., Nunes Filho, G., Jobert, A., et al. (2010). How learning to read changes the cortical networks for vision and language. *Science* 330, 1359–1364. doi: 10.1126/science.1194140
- Denckla, M. B., and Rudel, R. G. (1976). Rapid "automatized" naming (R.A.N.): dyslexia differentiated from other learning disabilities. *Neuropsychologia* 14, 471–479. doi: 10.1016/0028-3932(76)90075-0
- Desikan, R. S., Ségonne, F., Fischl, B., Quinn, B. T., Dickerson, B. C., Blacker, D., et al. (2006). An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *Neuroimage* 31, 968–980. doi: 10.1016/j.neuroimage.2006.01.021
- Desroches, A. S., Cone, N. E., Bolger, D. J., Bitan, T., Burman, D. D., and Booth, J. R. (2010). Children with reading difficulties show differences in brain regions associated with orthographic processing during spoken language processing. *Brain Res.* 1356, 73–84. doi: 10.1016/j.brainres.2010.07.097
- Eklund, K., Torppa, M., Aro, M., Leppänen, P. H. T., and Lyytinen, H. (2015). Literacy skill development of children with familial risk for dyslexia through grades 2, 3, and 8. *J. Educ. Psychol.* 107, 126–140. doi: 10.1037/a0037121
- Espy, K. A., Molfese, D. L., Molfese, V. J., and Modglin, A. (2004). Development of auditory event-related potentials in young children and relations to word-level reading abilities at age 8 years. *Ann. Dyslexia* 54, 9–38. doi: 10.1007/s11881-004-0002-3
- Fraga González, G., Žarić, G., Tijms, J., Bonte, M., Blomert, L., Leppänen, P., et al. (2016). Responsivity to dyslexia training indexed by the N170 amplitude of the brain potential elicited by word reading. *Brain Cogn.* 106, 42–54. doi: 10.1016/j.bandc.2016.05.001
- Fraga González, G., Žarić, G., Tijms, J., Bonte, M., Blomert, L., and van der Molen, M. W. (2014). Brain-potential analysis of visual word recognition in dyslexics and typically reading children. *Front. Hum. Neurosci.* 8:474. doi: 10.3389/fnhum.2014.00474
- Fraga González, G., Žarić, G., Tijms, J., Bonte, M., and van der Molen, M. W. (2017). Contributions of letter-speech sound learning and visual print tuning to reading improvement: evidence from brain potential and dyslexia training studies. *Brain Sci.* 7:E10. doi: 10.3390/brainsci7010010
- Froyen, D., van Atteveldt, N., and Blomert, L. (2010). Exploring the role of low level visual processing in letter-speech sound integration: a visual MMN study. *Front. Integr. Neurosci.* 4:9. doi: 10.3389/fnint.2010.00009
- Froyen, D., Van Atteveldt, N., Bonte, M., and Blomert, L. (2008). Cross-modal enhancement of the MMN to speech-sounds indicates early and automatic integration of letters and speech-sounds. *Neurosci. Lett.* 430, 23–28. doi: 10.1016/j.neulet.2007.10.014
- Froyen, D. J. W., Bonte, M. L., van Atteveldt, N., and Blomert, L. (2009). The long road to automation: neurocognitive development of letter-speech sound processing. *J. Cogn. Neurosci.* 21, 567–580. doi: 10.1162/jocn.2009.21061
- Gramfort, A., Luessi, M., Larson, E., Engemann, D. A., Strohmeier, D., Brodbeck, C., et al. (2013). MEG and EEG data analysis with MNE-Python. *Front. Neurosci.* 7:267. doi: 10.3389/fnins.2013.00267
- Groppe, D. M., Urbach, T. P., and Kutas, M. (2011). Mass univariate analysis of event-related brain potentials/fields I: a critical tutorial review. *Psychophysiology* 48, 1711–1725. doi: 10.1111/j.1469-8986.2011.01273.x
- Hämäläinen, J. A., Guttorm, T. K., Richardson, U., Alku, P., Lyytinen, H., and Leppänen, P. H. T. (2013). Auditory event-related potentials measured in kindergarten predict later reading problems at school age. *Dev. Neuropsychol.* 38, 550–566. doi: 10.1080/87565641.2012.718817
- Hämäläinen, J. A., Leppänen, P. H. T., Eklund, K., Thomson, J., Richardson, U., Guttorm, T. K., et al. (2009). Common variance in amplitude envelope perception tasks and their impact on phoneme duration perception and reading and spelling in Finnish children with reading disabilities. *Appl. Psycholinguist.* 30, 511–530. doi: 10.1017/s0142716409090250
- Hämäläinen, J. A., Lohvansuu, K., Ervast, L., and Leppänen, P. H. T. (2015). Event-related potentials to tones show differences between children with multiple risk factors for dyslexia and control children before the onset of formal reading instruction. *Int. J. Psychophysiol.* 95, 101–112. doi: 10.1016/j.ijpsycho.2014.04.004

- Hämäläinen, M. S., and Ilmoniemi, R. J. (1994). Interpreting magnetic fields of the brain: minimum norm estimates. *Med. Biol. Eng. Comput.* 32, 35–42. doi: 10.1007/BF02512476
- Hardy, M., Smythe, P. C., and Stennett, R. G. (1972). Developmental patterns in elemental reading skills: phoneme-grapheme and grapheme-phoneme correspondences. *J. Educ. Psychol.* 63, 433–436. doi: 10.1037/h0033240
- Häyriäinen, T., Serenius-Sirve, S., and Korkman, M. (1999). *Lukilasse. Lukemisen, Kirjoittamisen Ja Laskemisen Seulontatestistö Peruskoulun Ala-Asteen Luokille*. Helsinki: Psykologien Kustannus Oy, 1–6.
- Hein, G., Doehrmann, O., Müller, N. G., Kaiser, J., Muckli, L., and Naumer, M. J. (2007). Object familiarity and semantic congruency modulate responses in cortical audiovisual integration areas. *J. Neurosci.* 27, 7881–7887. doi: 10.1523/JNEUROSCI.1740-07.2007
- Herdman, A. T., Fujioka, T., Chau, W., Ross, B., Pantev, C., and Picton, T. W. (2006). Cortical oscillations related to processing congruent and incongruent grapheme-phoneme pairs. *Neurosci. Lett.* 399, 61–66. doi: 10.1016/j.neulet.2006.01.069
- Hyvärinen, A., and Oja, E. (2000). Independent component analysis: algorithms and applications. *Neural Netw.* 13, 411–430. doi: 10.1016/s0893-6080(00)00026-5
- Jas, M., Larson, E., Engemann, D.-A., Leppakangas, J., Taulu, S., Hamalainen, M., et al. (2017). MEG/EEG group study with MNE: recommendations, quality assessments and best practices. *bioRxiv* [Preprint]. doi: 10.1101/240044
- Jones, J. A., and Callan, D. E. (2003). Brain activity during audiovisual speech perception: an fMRI study of the McGurk effect. *Neuroreport* 14, 1129–1133. doi: 10.1097/01.wnr.0000074343.81633.2a
- Jost, L. B., Eberhard-Moscicka, A. K., Frisch, C., Dellwo, V., and Maurer, U. (2014). Integration of spoken and written words in beginning readers: a topographic ERP study. *Brain Topogr.* 27, 786–800. doi: 10.1007/s10548-013-0336-4
- Katzir, T., Misra, M., and Poldrack, R. A. (2005). Imaging phonology without print: assessing the neural correlates of phonemic awareness using fMRI. *Neuroimage* 27, 106–115. doi: 10.1016/j.neuroimage.2005.04.013
- Korkman, M., Kirk, U., and Kemp, S. (2007). *NEPSY (NEPSY-II)*, 2nd Edn. San Antonio, TX: Harcourt Assessment.
- Kushnerenko, E., Ceponiene, R., Balan, P., Fellman, V., Huotilaine, M., and Näätänen, R. (2002). Maturation of the auditory event-related potentials during the first year of life. *Neuroreport* 13, 47–51. doi: 10.1097/00001756-200201210-00014
- Kuuluvainen, S., Leminen, A., and Kujala, T. (2016). Auditory evoked potentials to speech and nonspeech stimuli are associated with verbal skills in preschoolers. *Dev. Cogn. Neurosci.* 19, 223–232. doi: 10.1016/j.dcn.2016.04.001
- Lin, F. H., Witzel, T., Ahlfors, S. P., Stufflebeam, S. M., Belliveau, J. W., and Hämäläinen, M. S. (2006). Assessing and improving the spatial accuracy in MEG source localization by depth-weighted minimum-norm estimates. *Neuroimage* 31, 160–171. doi: 10.1016/j.neuroimage.2005.11.054
- Lohvansuu, K., Hämäläinen, J. A., Ervast, L., Lyytinen, H., and Leppänen, P. H. T. (2018). Longitudinal interactions between brain and cognitive measures on reading development from 6 months to 14 years. *Neuropsychologia* 108, 6–12. doi: 10.1016/j.neuropsychologia.2017.11.018
- Lohvansuu, K., Hämäläinen, J. A., Tanskanen, A., Ervast, L., Heikkinen, E., Lyytinen, H., et al. (2014). Enhancement of brain event-related potentials to speech sounds is associated with compensated reading skills in dyslexic children with familial risk for dyslexia. *Int. J. Psychophysiol.* 94, 298–310. doi: 10.1016/j.ijpsycho.2014.10.002
- Lovio, R., Näätänen, R., and Kujala, T. (2010). Abnormal pattern of cortical speech feature discrimination in 6-year-old children at risk for dyslexia. *Brain Res.* 1335, 53–62. doi: 10.1016/j.brainres.2010.03.097
- Lyytinen, H., Erskine, J., Kujala, J., Ojanen, E., and Richardson, U. (2009). In search of a science-based application: a learning tool for reading acquisition. *Scand. J. Psychol.* 50, 668–675. doi: 10.1111/j.1467-9450.2009.00791.x
- Maris, E., and Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *J. Neurosci. Methods* 164, 177–190. doi: 10.1016/j.jneumeth.2007.03.024
- Maurer, U., Blau, V. C., Yoncheva, Y. N., and McCandliss, B. D. (2010). Development of visual expertise for reading: rapid emergence of visual familiarity for an artificial script. *Dev. Neuropsychol.* 35, 404–422. doi: 10.1080/87565641.2010.480916
- Maurer, U., Brem, S., Bucher, K., and Brandeis, D. (2005). Emerging neurophysiological specialization for letter strings. *J. Cogn. Neurosci.* 17, 1532–1552. doi: 10.1162/089892905774597218
- Maurer, U., Brem, S., Kranz, F., Bucher, K., Benz, R., Halder, P., et al. (2006). Coarse neural tuning for print peaks when children learn to read. *Neuroimage* 33, 749–758. doi: 10.1016/j.neuroimage.2006.06.025
- Maurer, U., Zevin, J. D., and McCandliss, B. D. (2008). Left-lateralized N170 effects of visual expertise in reading: evidence from Japanese syllabic and logographic scripts. *J. Cogn. Neurosci.* 20, 1878–1891. doi: 10.1162/jocn.2008.20125
- McCandliss, B. D., Cohen, L., and Dehaene, S. (2003). The visual word form area: expertise for reading in the fusiform gyrus. *Trends Cogn. Sci.* 7, 293–299. doi: 10.1016/S1364-6613(03)00134-7
- McCandliss, B. D., and Noble, K. G. (2003). The development of reading impairment: a cognitive neuroscience model. *Ment. Retard. Dev. Disabil. Res. Rev.* 9, 196–204. doi: 10.1002/mrdd.10080
- Melby-Lervåg, M., Lyster, S.-A. H., and Hulme, C. (2012). Phonological skills and their role in learning to read: a meta-analytic review. *Psychol. Bull.* 138, 322–352. doi: 10.1037/a0026744
- Näätänen, R. (2001). The perception of speech sounds by the human brain as reflected by the mismatch negativity (MMN) and its magnetic equivalent (MMNm). *Psychophysiology* 38, 1–21. doi: 10.1111/1469-8986.3810001
- Ojanen, V., Möttönen, R., Pekkola, J., Jääskeläinen, I. P., Joensuu, R., Autti, T., et al. (2005). Processing of audiovisual speech in Broca's area. *Neuroimage* 25, 333–338. doi: 10.1016/j.neuroimage.2004.12.001
- Parviainen, T., Helenius, P., Poskiparta, E., Niemi, P., and Salmelin, R. (2006). Cortical sequence of word perception in beginning readers. *J. Neurosci.* 26, 6052–6061. doi: 10.1523/JNEUROSCI.0673-06.2006
- Parviainen, T., Helenius, P., Poskiparta, E., Niemi, P., and Salmelin, R. (2011). Speech perception in the child brain: cortical timing and its relevance to literacy acquisition. *Hum. Brain Mapp.* 32, 2193–2206. doi: 10.1002/hbm.21181
- Parviainen, T., Helenius, P., and Salmelin, R. (2005). Cortical differentiation of speech and nonspeech sounds at 100 ms: implications for dyslexia. *Cereb. Cortex* 15, 1054–1063. doi: 10.1093/cercor/bhh206
- Pennington, B. F., and Lefly, D. L. (2001). Early reading development in children at family risk for dyslexia. *Child Dev.* 72, 816–833. doi: 10.1111/1467-8624.00317
- Ponton, C. W., Eggermont, J. J., Kwong, B., and Don, M. (2000). Maturation of human central auditory system activity: evidence from multi-channel evoked potentials. *Clin. Neurophysiol.* 111, 220–236. doi: 10.1016/S1388-2457(99)00236-9
- Preston, J. L., Molfese, P. J., Frost, S. J., Mencl, W. E., Fulbright, R. K., Hoef, F., et al. (2016). Print-speech convergence predicts future reading outcomes in early readers. *Psychol. Sci.* 27, 75–84. doi: 10.1177/0956797615611921
- Price, C. J. (2000). The anatomy of language: contributions from functional neuroimaging. *J. Anat.* 197, 335–359. doi: 10.1046/j.1469-7580.2000.19730335.x
- Pugh, K. R., Landi, N., Preston, J. L., Mencl, W. E., Austin, A. C., Sibley, D., et al. (2013). The relationship between phonological and auditory processing and brain organization in beginning readers. *Brain Lang.* 125, 173–183. doi: 10.1016/j.bandl.2012.04.004
- Pugh, K. R., Mencl, W. E., Jenner, A. R., Katz, L., Frost, S. J., Lee, J. R., et al. (2000a). Functional neuroimaging studies of reading and reading disability (developmental dyslexia). *Ment. Retard. Dev. Disabil. Res. Rev.* 6, 207–213. doi: 10.1002/1098-2779(2000)6:3<207::AID-MRDD8>3.0.CO;2-P
- Pugh, K. R., Mencl, W. E., Shaywitz, B. A., Shaywitz, S. E., Fulbright, R. K., Constable, R. T., et al. (2000b). The angular gyrus in developmental dyslexia: task-specific differences in functional connectivity within posterior cortex. *Psychol. Sci.* 11, 51–56. doi: 10.1111/1467-9280.00214
- Pugh, K. R., Mencl, W. E., Jenner, A. R., Katz, L., Frost, S. J., Lee, J. R., et al. (2001). Neurobiological studies of reading and reading disability. *J. Commun. Disord.* 34, 479–492. doi: 10.1016/S0021-9924(01)00060-0
- Puolakanaho, A., Ahonen, T., Aro, M., Eklund, K., Leppänen, P. H. T., Poikkeus, A.-M., et al. (2007). Very early phonological and language skills: estimating individual risk of reading disability. *J. Child Psychol. Psychiatry* 48, 923–931. doi: 10.1111/j.1469-7610.2007.01763.x
- Raij, T., Uutela, K., and Hari, R. (2000). Audiovisual integration of letters in the human brain. *Neuron* 28, 617–625. doi: 10.1016/S0896-6273(00)00138-0

- Raschle, N. M., Chang, M., and Gaab, N. (2011). Structural brain alterations associated with dyslexia predate reading onset. *Neuroimage* 57, 742–749. doi: 10.1016/j.neuroimage.2010.09.055
- Raschle, N. M., Zuk, J., and Gaab, N. (2012). Functional characteristics of developmental dyslexia in left-hemispheric posterior brain regions predate reading onset. *Proc. Natl. Acad. Sci. U.S.A.* 109, 2156–2161. doi: 10.1073/pnas.1107721109
- Rueckl, J. G., Paz-Alonso, P. M., Molfese, P. J., Kuo, W.-J., Bick, A., Frost, S. J., et al. (2015). Universal brain signature of proficient reading: evidence from four contrasting languages. *Proc. Natl. Acad. Sci. U.S.A.* 112, 15510–15515. doi: 10.1073/pnas.1509321112
- Rüsseler, J., Ye, Z., Gerth, I., Szycik, G. R., and Münte, T. F. (2017). Audio-visual speech perception in adult readers with dyslexia: an fMRI study. *Brain Imaging Behav.* 12, 357–368. doi: 10.1007/s11682-017-9694-y
- Sandak, R., Einar Mencl, W., Frost, S. J., and Pugh, K. R. (2004). The neurobiological basis of skilled and impaired reading: recent findings and new directions. *Sci. Stud. Read.* 8, 273–292. doi: 10.1207/s1532799xssr0803_6
- Schlaggar, B. L., and McCandliss, B. D. (2007). Development of neural systems for reading. *Annu. Rev. Neurosci.* 30, 475–503. doi: 10.1146/annurev.neuro.28.061604.135645
- Shankweiler, D., Mencl, W. E., Braze, D., Tabor, W., Pugh, K. R., and Fulbright, R. K. (2008). Reading differences and brain: cortical integration of speech and print in sentence processing varies with reader skill. *Dev. Neuropsychol.* 33, 745–775. doi: 10.1080/87565640802418688
- Sliwiska, M. W., James, A., and Devlin, J. T. (2015). Inferior parietal lobe contributions to visual word recognition. *J. Cogn. Neurosci.* 27, 593–604. doi: 10.1162/jocn_a_00721
- Specht, K., Hugdahl, K., Ofte, S., Nygård, M., Bjørnerud, A., Plante, E., et al. (2009). Brain activation on pre-reading tasks reveals at-risk status for dyslexia in 6-year-old children. *Scand. J. Psychol.* 50, 79–91. doi: 10.1111/j.1467-9450.2008.00688.x
- Sperdin, H. F., Cappe, C., Foxe, J. J., and Murray, M. M. (2009). Early, low-level auditory-somatosensory multisensory interactions impact reaction time speed. *Front. Integr. Neurosci.* 3:2. doi: 10.3389/neuro.07.002.2009
- Stein, B. E., and Stanford, T. R. (2008). Multisensory integration: current issues from the perspective of the single neuron. *Nat. Rev. Neurosci.* 9, 255–266. doi: 10.1038/nrn2331
- Stevens, C., McIlraith, A., Rusk, N., Niermeyer, M., and Waller, H. (2013). Relative laterality of the N170 to single letter stimuli is predicted by a concurrent neural index of implicit processing of letternames. *Neuropsychologia* 51, 667–674. doi: 10.1016/j.neuropsychologia.2012.12.009
- Szymanski, M. D., Rowley, H. A., and Roberts, T. P. (1999). A hemispherically asymmetrical MEG response to vowels. *Neuroreport* 10, 2481–2486. doi: 10.1097/00001756-199908200-00009
- Taulu, S., and Kajola, M. (2005). Presentation of electromagnetic multichannel data: the signal space separation method. *J. Appl. Phys.* 97:124905. doi: 10.1063/1.1935742
- Taulu, S., Kajola, M., and Simola, J. (2004). Suppression of interference and artifacts by the signal space separation method. *Brain Topogr.* 16, 269–275. doi: 10.1023/B:BRAT.0000032864.93890.f9
- Taulu, S., and Simola, J. (2006). Spatiotemporal signal space separation method for rejecting nearby interference in MEG measurements. *Phys. Med. Biol.* 51, 1759–1768. doi: 10.1088/0031-9155/51/7/008
- Torgesen, J. K., Rashotte, C. A., and Wagner, R. K. (1999). *TOWRE: Test of Word Reading Efficiency*. Austin, TX: Pro-Ed.
- Tremblay, C., Champoux, F., Voss, P., Bacon, B. A., Lepore, F., and Théoret, H. (2007). Speech and non-speech audio-visual illusions: a developmental study. *PLoS One* 2:e742. doi: 10.1371/journal.pone.0000742
- van Atteveldt, N., Formisano, E., Goebel, R., and Blomert, L. (2004). Integration of letters and speech sounds in the human brain. *Neuron* 43, 271–282. doi: 10.1016/j.neuron.2004.06.025
- van Atteveldt, N., Roebroek, A., and Goebel, R. (2009). Interaction of speech and script in human auditory cortex: insights from neuro-imaging and effective connectivity. *Hear. Res.* 258, 152–164. doi: 10.1016/j.heares.2009.05.007
- van Atteveldt, N. M., Formisano, E., Goebel, R., and Blomert, L. (2007). Top-down task effects overrule automatic multisensory responses to letter-sound pairs in auditory association cortex. *Neuroimage* 36, 1345–1360. doi: 10.1016/j.neuroimage.2007.03.065
- Vandermosten, M., Hoef, F., and Norton, E. S. (2016). Integrating MRI brain imaging studies of pre-reading children with current theories of developmental dyslexia: a review and quantitative meta-analysis. *Curr. Opin. Behav. Sci.* 10, 155–161. doi: 10.1016/j.cobeha.2016.06.007
- Weber-Fox, C., Spencer, R., Cuadrado, E., and Smith, A. (2003). Development of neural processes mediating rhyme judgments: phonological and orthographic interactions. *Dev. Psychobiol.* 43, 128–145. doi: 10.1002/dev.10128
- Wechsler, D. (1991). *Manual for the Wechsler Intelligence Scale for Children-(WISC-III)*. San Antonio, TX: Psychological Corporation.
- Wechsler, D. (2003). *Wechsler Intelligence Scale for Children*, 4th Edn. San Antonio, TX: The Psychological Corporation.
- Wunderlich, J. L., Cone-Wesson, B. K., and Shepherd, R. (2006). Maturation of the cortical auditory evoked potential in infants and young children. *Hear. Res.* 212, 185–202. doi: 10.1016/j.heares.2005.11.010
- Yoncheva, Y. N., Zevin, J. D., Maurer, U., and McCandliss, B. D. (2010). Auditory selective attention to speech modulates activity in the visual word form area. *Cereb. Cortex* 20, 622–632. doi: 10.1093/cercor/bhp129
- Žarić, G., González, G. F., Tijms, J., Van Der Molen, M. W., Blomert, L., and Bonte, M. (2014). Reduced neural integration of letters and speech sounds in dyslexic children scales with individual differences in reading fluency. *PLoS One* 9:e110337. doi: 10.1371/journal.pone.0110337

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Xu, Kolozsvari, Monto and Hämäläinen. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



III

RAPID CHANGES IN BRAIN ACTIVITY DURING LEARNING OF GRAPHEME-PHONEME ASSOCIATIONS IN ADULTS

by

Weiyong Xu, Orsolya Beatrix Kolozsvári, Robert Oostenveld and Jarmo Arvid
Hämäläinen, 2020

NeuroImage, 117058

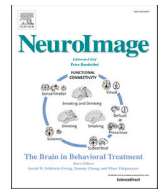
<https://doi.org/10.1016/j.neuroimage.2020.117058>

Reproduced with kind permission by Elsevier.



Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/neuroimage

Rapid changes in brain activity during learning of grapheme-phoneme associations in adults



Weiyong Xu^{a,b,*}, Orsolya Beatrix Kolozsvari^{a,b}, Robert Oostenveld^{c,d}, Jarmo Arvid Hämäläinen^{a,b}

^a Department of Psychology, University of Jyväskylä, Jyväskylä, Finland

^b Jyväskylä Centre for Interdisciplinary Brain Research, University of Jyväskylä, Jyväskylä, Finland

^c Donders Institute for Brain, Cognition and Behaviour, Radboud University, Nijmegen, Netherlands

^d NatMEG, Department of Clinical Neuroscience, Karolinska Institutet, Stockholm, Sweden

ARTICLE INFO

Keywords:

Reading
Audiovisual
Learning
Magnetoencephalography

ABSTRACT

Learning to associate written letters with speech sounds is crucial for the initial phase of acquiring reading skills. However, little is known about the cortical reorganization for supporting letter-speech sound learning, particularly the brain dynamics during the learning of grapheme-phoneme associations. In the present study, we trained 30 Finnish participants (mean age: 24.33 years, SD: 3.50 years) to associate novel foreign letters with familiar Finnish speech sounds on two consecutive days (first day ~ 50 min; second day ~ 25 min), while neural activity was measured using magnetoencephalography (MEG). Two sets of audiovisual stimuli were used for the training in which the grapheme-phoneme association in one set (Learnable) could be learned based on the different learning cues provided, but not in the other set (Control). The learning progress was tracked at a trial-by-trial basis and used to segment different learning stages for the MEG source analysis. The learning-related changes were examined by comparing the brain responses to Learnable and Control uni/multi-sensory stimuli, as well as the brain responses to learning cues at different learning stages over the two days. We found dynamic changes in brain responses related to multi-sensory processing when grapheme-phoneme associations were learned. Further, changes were observed in the brain responses to the novel letters during the learning process. We also found that some of these learning effects were observed only after memory consolidation the following day. Overall, the learning process modulated the activity in a large network of brain regions, including the superior temporal cortex and the dorsal (parietal) pathway. Most interestingly, middle- and inferior-temporal regions were engaged during multi-sensory memory encoding after the cross-modal relationship was extracted from the learning cues. Our findings highlight the brain dynamics and plasticity related to the learning of letter-speech sound associations and provide a more refined model of grapheme-phoneme learning in reading acquisition.

1. Introduction

The learning of grapheme-phoneme associations is a crucial step for reading acquisition in alphabetic languages. Unlike spoken language, written script is a recent cultural invention, and therefore no hard-wired brain circuit exists for reading at birth (Lieberman, 1992; Lieberman, 2006). Consequently, reading acquisition involves plastic changes in pre-existing structural and functional networks of the brain, such as the visual and language systems to meet the new cognitive demand in

reading (Dehaene et al. 2010, 2015). However, little is known about the cognitive processes and neural systems engaged during the learning of letter-sound correspondences, since most existing studies examine the long-term effects of learning to read.

One of the most important brain-level markers in learning to read is the increasing sensitivity to orthographic stimuli in the left ventral occipitotemporal cortex (vOT) (Brem et al., 2010; Ben-Shachar et al., 2011; Dehaene et al., 2010; Dehaene and Cohen, 2011). The left vOT connects the visual word forms to other language areas of the brain, and

* Corresponding author. Jyväskylä Centre for Interdisciplinary Brain Research, Department of Psychology, University of Jyväskylä, P.O. Box 35, 40014, University of Jyväskylä, Finland.

E-mail addresses: weiyong.w.xu@jyu.fi (W. Xu), orsolya.b.kolozsvari@jyu.fi (O.B. Kolozsvari), r.oostenveld@donders.ru.nl (R. Oostenveld), jarmo.a.hamalainen@jyu.fi (J.A. Hämäläinen).

<https://doi.org/10.1016/j.neuroimage.2020.117058>

Received 24 March 2020; Received in revised form 11 June 2020; Accepted 12 June 2020

Available online 17 June 2020

1053-8119/© 2020 The Author(s). Published by Elsevier Inc. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

has a posterior to anterior gradient (Lerma-Usabiaga et al., 2018; Vinckier et al., 2007), with the posterior part responsible for visual feature extraction and sensitive to smaller grain sizes (e.g., letters) and the anterior part to larger grain sizes (e.g., words (Dehaene et al., 2005)). Furthermore, vOT interacts with spoken language systems, for example, the phonological representations in the temporal cortex (Price and Devlin, 2011). The development of letter specificity (greater activation for letters compared with false fonts) in left vOT is associated with reading ability in beginning readers (Centanni et al., 2018).

While changes in brain activity in the vOT is a hallmark for reading acquisition, learning to read requires other additional processes. A large body of neuroimaging studies (van Atteveldt et al., 2009; Raij et al., 2000; Blau et al., 2008) has looked into the audiovisual integration of graphemes and phonemes in literate adults who have mastered the associations through initial learning in childhood and years of the reading experience afterward. Consistent findings (Beauchamp et al., 2004a; van Atteveldt et al., 2004; Richlan, 2019; Blau et al., 2008; Calvert, 2001; Wilson et al., 2018) suggest that the multisensory superior temporal cortex (STC) is the major brain region for audiovisual integration. For example, the brain activation to congruent audiovisual stimuli (letter-speech sound combinations) was found to be stronger than incongruent audiovisual stimuli in the left STC in transparent orthographies such as Dutch (van Atteveldt et al., 2004) and Finnish (Raij et al., 2000). In addition, the audiovisual congruency effect seems to be modulated by the transparency of the orthography as well as the task requirements used during the experiments (van Atteveldt et al., 2007). For example, audiovisual incongruent stimuli elicit larger neural responses than congruent stimuli in more opaque orthographies such as English (Hollaway et al., 2015) and logographic scripts such as Chinese (Xu et al., 2019). The level of automaticity in audiovisual integration is important for normal reading development (Xu et al., 2018; Varga et al., 2020) and failing of which has been shown to be closely linked to dyslexia (Blomert, 2011; Žarić et al., 2014; Blau et al., 2010). Automaticity for letter-speech correspondences seems to develop slowly, with electrophysiological mismatch negativity (MMN) studies showing a prolonged trajectory of audiovisual integration in children for up to 4 years after reading acquisition (Froyen et al., 2009). The slow development of automaticity is partly related to the neural representation of learned letters that must account for the numerous variations in position, case, and font for the same letter. This poses a great demand for reorganizing the hierarchical letter processing pathway along the vOT regions (Dehaene et al., 2005).

Brain processes for grapheme-phoneme associations have been studied for well-established associations in literate adults (Raij et al., 2000; van Atteveldt et al., 2004, 2007, 2009; Blau et al., 2008; Froyen et al., 2008) and children at different stages of learning to read (Žarić et al., 2014; Froyen et al., 2009, 2011; Blau et al., 2010). These studies showed brain networks that are consistently activated during audiovisual integration days, months, or even years after learning of grapheme-phoneme associations. Less is known about the cognitive processes during the learning of new associations, which is arguably more complex and demanding than the automatic processing of existing associations. The scarcity of cross-modal studies on the learning process in humans is likely due to challenges in studying the brain mechanism during multisensory learning since it is very dynamic and involves multiple cognitive components such as sensory processing, multisensory integration, attention, memory formation, and consolidation.

The grapheme-phoneme learning process likely consists of several stages: First, during explicit learning, attention is directed to the information coming from the auditory and visual modalities. In addition, auditory and visual stimuli are combined into audiovisual objects in multisensory brain regions (Stein and Stanford, 2008) (e.g., STC) and such cross-modal audiovisual association is initially stored in the short-term memory system. The short-term memories of audiovisual associations are consolidated during both practice and sleep (Diekelmann and Born, 2010; Dudai, 2012). They are then most likely transferred and stored in the neocortex for fast and automatic retrieval (Klinzing et al.,

2019). Complementary learning systems have been shown in the medial-temporal systems (hippocampus and parahippocampal cortex) and neocortex, and a division of labor with the initial rapid learning in the hippocampus and gradual memory consolidation in the neocortical systems (McClelland et al., 1995; Davis et al., 2009). Converging evidence from a recent neuroimaging study also suggests a crucial role of the parahippocampal cortex for symbolic learning (Skeide et al., 2018). However, fast learning effects that occurred as a rapid form of memory consolidation at the time scale of seconds have also been reported in relation to motor-skill learning (Bönstrup et al., 2019). Such rapid consolidation might also play a role in other types of sensory learning (Hebscher et al., 2019).

Artificial grapheme-phoneme training paradigms that simulate the initial stage of learning to read in alphabetic scripts have provided interesting insights into the brain mechanisms of learning grapheme-phoneme associations. Learning-related brain changes have been reported at the time scale of minutes (Hämäläinen et al., 2019; Karipidis et al., 2017), hours (Taylor et al., 2014; Brem et al., 2018) and days (Taylor et al., 2017; Hashimoto and Sakai, 2004; Madec et al., 2016; Quinn et al., 2017; Karipidis et al., 2018) after initial training. Combining information from visual and auditory modalities has been suggested to involve at least two possible mechanisms: Hashimoto and Sakai reported the involvement of the left posterior inferior temporal gyrus (PITG) and the left parieto-occipital cortex (PO) which showed plasticity for forming new links between orthography and phonology when learning novel letters (Hashimoto and Sakai, 2004). The involvement of parietal brain regions has also been reported to be crucial for exponential visual-verbal mappings in the early stages of learning to read (Taylor et al., 2014; Quinn et al., 2017). On the other hand (Madec et al., 2016), have shown left vOT to be involved in phonological recoding processes of newly-learned letters by top-down influences from STG; this effect was affected by the strength of audiovisual associations in a two-day letter-speech sound training. Similar changes in the left vOT have been reported to show larger N170 responses (Brem et al., 2018) and vOT activation to trained than untrained characters after a short artificial character-speech sound training. These changes were also correlated with the training performance and were interpreted as a phonologically driven N170 and vOT tuning (Pleisch et al., 2019). Furthermore, these processes might be affected by modulation of attention to important features for learning in the frontal cortices (Hämäläinen et al., 2019). Interestingly the fast learning-related changes in brain activity seem to be linked to cognitive performance (Karipidis et al., 2017, 2018). Multisensory integration effects were found in a distributed brain network after a short letter-speech sound training (<30 min) in preschool children (Karipidis et al., 2017) with promising implications for identifying poor-reading children and predicting reading outcomes in pre-readers (Karipidis et al., 2018).

Despite the emerging insights from the available literature, to date, there is no comprehensive theoretical model of the cognitive processes and their brain level equivalents that are utilized during grapheme-phoneme learning. It is unclear when and how the audiovisual congruency effect starts to emerge in the multisensory superior temporal cortex and how quickly during training the visual representation of learned letters starts to differ from unfamiliar letters. Also, the allocation of attention is essential during explicit learning, yet how attentional processing is modulated by the learning material is still unknown. Finally, brain changes during the early stages of memory consolidation are still poorly understood, for example, effects related to repetition and practice during the initial learning stage, and the effect of overnight sleep on memory consolidation of letter-speech sound associations.

In our study, we investigated the neural mechanisms during the learning of novel grapheme-phoneme associations and the effect of overnight memory consolidation. The learning progress was tracked at a trial-by-trial basis during training on two consecutive days and was used to identify and segment different learning stages. Learning effects specifically related to grapheme-phoneme associations were studied by

using two audiovisual stimulus sets. In one set, the audiovisual associations could be learned (cross-modal associative learning); in the other set, this was not possible due to the absence of information on the correct cross-modal correspondences (cross-modal non-associative learning). Different learning cues were presented after the audiovisual stimuli to dissociate the learning of correct audiovisual associations from basic multi-sensory processes. During cross-modal associative learning, the auditory and visual inputs had to be integrated and encoded into one audiovisual object, while no such integrative processes were needed in non-associative learning. We expected to see distinctive cognitive processes related to attention and memory encoding in non-associative and associative learning. Furthermore, we hypothesized that the learning of grapheme-phoneme associations would change the corresponding unisensory visual processing and elicit multisensory congruency effects. The unisensory effects were expected to occur in occipital and parietal regions due to the learning of the phonological representation of the Learnable letters mostly at a relatively late time window around 400 ms based on earlier studies (Xu et al. 2018, 2019; Quinn et al., 2017; Taylor et al., 2014; Dehaene et al., 2010). The multisensory congruency effects were expected to be elicited in the posterior superior temporal cortices in the late time window only after the learning of audiovisual associations (Wilson et al., 2018; van Atteveldt et al., 2004; Xu et al., 2019). These cross-modal learning effects were expected to be modulated by overnight memory consolidation. Finally, learning performance was correlated with cognitive skills linked to reading and working memory to explore the key behavioral factors that contribute to multisensory non-associative/associative learning speed.

2. Material and methods

2.1. Participants

In total, 36 people were scheduled to participate in the study. The participants were university students and staff recruited through email-lists and posters. The data from 6 participants were not included: 2 were excluded due to a low learning accuracy during the whole training session on Day 1, and for the other 4 participants, the MEG data were not measured because of cancellation. Therefore the data from the remaining thirty participants (20 females; 26 right-handed, 2 ambidextrous; mean age 24.3 years, SD 3.5 years, range 19–36 years) were used in this study. All participants had normal hearing and normal or corrected-to-normal vision (based on self-report). Participants were screened for the following exclusion criteria: head injuries, ADHD, neurological diseases, medication affecting the central nervous system, delays in language development, or any other language-related disorders. The ethical approval for this study was obtained from the Ethics Committee of the University of Jyväskylä, and the study was carried out in accordance with the Declaration of Helsinki. Participants gave their written informed consent prior to their participation in the experiments. After the MEG experiments, all of them received movie tickets or gift cards (for an equivalent value of 30 euros) as compensation for their time in the MEG recording and cognitive test sessions (see details of cognitive tests below).

2.2. Stimuli and task

Visual stimuli consisted of 12 Georgian letters (ჲ, ჳ, ჴ, ჵ, ჶ, ჷ, ჸ, ჹ, ჺ, ჻, ჼ, ჽ, ჾ, ჿ). Auditory stimuli consisted of 12 Finnish phonemes ([a], [ä], [e], [t], [s], [k], [o], [ö], [i], [p], [v], [d]; mean duration: 473 ms; SD:103 ms). The auditory and visual stimuli were divided into two sets with 6 audiovisual pairs in each set. The stimuli between set one (ჲ: [a], ჳ: [ä], ჴ: [e], ჵ: [t], ჶ: [s], ჷ: [k]) and set two (ღ: [o], ჹ: [ö], ჺ: [i], ჻: [p], ჼ: [v], ჽ: [d]) were chosen to match as closely as possible in visual and auditory complexity. The auditory stimuli in each of the Learnable and Control sets included three vowels and three consonants. The types of the phonemes were counterbalanced between the two sets (vowels were picked

in the closed-open dimension: [e], [ä], [a] and [i], [ö], [o]; stop consonants: [p], [d], [t], [k]; fricative consonants: [s], [v]). The visual letters in the Learnable and Control sets were roughly matched in the overall shape and curvatures. Four additional letter-sound pairs (ო: [y], ჳ: [u], ჴ: [b], ჵ: [g]) were used only for experimental instruction and practise purpose at the beginning of the experiment.

The audiovisual learning experiment consisted of 12 alternating training and testing blocks on Day 1, and 6 training and testing blocks on Day 2. In the training block, one of the two audiovisual stimulus sets was used as the Learnable set in which different learning cues (✓ for congruent pairs (AVC) and X for incongruent pairs (AVI)) were presented after the simultaneous presentation of audiovisual stimuli. The other audiovisual stimuli set was used as the Control set, in which the feedback was always ✘ after the randomly-paired audiovisual stimuli (AVX). The audiovisual trial started with a fixation cross presented for 1000 ms, followed by the audiovisual stimuli for another 1000 ms. After the audiovisual stimuli, one of the three learning cues (“YES”: ✓; “NO”: X; “UNKNOWN”: ✘) was presented for 1500 ms, depending on the types of audiovisual stimulus. Learnable and Control audiovisual stimuli were mixed and presented randomly in each training and testing block. In addition, auditory-only and visual-only stimuli from both sets were also presented in the training block, randomly mixed with the audiovisual stimuli.

Each training block was followed by a testing block in which the audiovisual stimuli from the training block were presented in random order, followed by a question on the stimulus congruency. The participants saw the question (“Match?”) displayed in the upper part of the screen and had to choose from three options (“YES”: ✓; “NO”: X; “UNKNOWN”: ✘) provided on the lower part of the screen using a response pad. The order of the 3 options was randomized so that the participants would not learn to associate specific response buttons with specific options. After finishing the testing block, the feedback was provided about the accuracy of the previous block and all blocks they had done so far. This was followed by a break in which participants were instructed to press a button to start the next training block when they were ready. The Learnable and Control sets were counterbalanced between the participants. Instructions for the participants and a short practice were given prior to the actual experiment on Day 1. Fig. 1 shows the stimuli and experiment design of the study.

2.3. MEG recording

MEG data were collected using the Elekta Neuromag® TRIUXTM system (Elekta AB, Stockholm, Sweden) in a magnetically shielded room at the University of Jyväskylä. A sampling rate of 1000 Hz and an online band-pass filter of 0.1–330 Hz were used in the data acquisition settings. The participant’s head position with respect to the MEG sensor arrays in the helmet was tracked continuously with five digitized head position indicator (HPI) coils. Three HPI coils were placed on the forehead and one behind each ear. The MEG head coordinate system was defined by three anatomic landmarks (nasion, left and right preauricular points). The anatomical landmarks, the position of the HPI coils, and the head shape (>100 points evenly distributed over the scalp) were digitized using a Polhemus Isotrak digital tracker system (Polhemus, Colchester, VT, United States) before the MEG experiment. The electrooculogram (EOG) was recorded with two electrodes attached diagonally slightly below the left and slightly above the right eye and one ground electrode attached to the collarbone. The MEG was recorded in a 68° upright gantry position with participants sitting comfortably on a chair. The same preparation and setup were used on Day 2.

2.4. Cognitive tests

A number of cognitive tests were administered to the participants to ensure they did not have language-related learning problems. Additional behavioral tests were conducted to run correlational analyses between learning speed in the MEG task and cognitive skill levels. The behavioral

A

Stimuli	SET 1						SET 2					
Auditory	[a]	[ä]	[e]	[t]	[s]	[k]	[o]	[ö]	[i]	[p]	[v]	[d]
Visual	ა	ჱ	ე	თ	ს	კ	ო	ჲ	ი	პ	ვ	დ

B

Block	BLOCK 1								BLOCK 2				...			
Task	Training				Testing				Training		Testing		
Category	Learnable		Control		Learnable		Control	
Stimuli (randomized)	A	V	A	V	A	V	A	V	A	V	A	V	A	V	A	V
Learning cues/Test	✓	✗							Match?							
									1	✗	2	✗	3	✓		

Fig. 1. The experimental task on letter-speech sound association learning. A) Auditory stimuli consisted of 12 Finnish phonemes. Visual stimuli consisted of 12 Georgian letters. The auditory and visual stimuli were divided into two sets (counterbalanced between participants as Learnable or Control) with 6 audiovisual pairs in each set. B) The audiovisual learning experiment consisted of 12 alternating training and testing blocks on Day 1, and 6 training and testing blocks on Day 2. After presenting the audiovisual stimuli, different learning cues were provided in the training block. In the testing block, learning progress was tracked by asking questions on the audiovisual congruence, which the participants had to answer using a response pad.

tests included the following: block design (visuospatial reasoning), vocabulary (expressive vocabulary), and digit span (forward and backward; working memory) from the Wechsler Adult Intelligence Scales (Wechsler, 2008). In the block design test, the participants were shown how to arrange blocks with red and white colors to form a design, and they have to build the same design. In more difficult sections, the participants are only shown the design in a figure, and they have to build it. In the vocabulary test, the participants hear a word, and they have to describe the meaning of that word. In the digit span test, a series of numbers is said to the participant, and they have to repeat them either in a forward or backward order. The mean of the standardized scores in these tests was 10 and the standard deviation was 3.

Phonological awareness was tested using the Phonological processing task from NEPSY II (Korkman et al., 2007). In this task, the participant is asked to repeat a word and then to create a new word by leaving out a syllable or a phoneme, or by replacing one phoneme in the word with another phoneme.

Non-word repetition task from the Neuropsychological test battery (NEPSY; Korkman et al., 1998) was used to measure phonological processing and verbal short-term memory. The number of correct items out of 16 was used as the score.

Rapid automatized naming (Denckla and Rudel, 1976), in which pictures of five common objects or five letters had to be named as quickly and as accurately as possible. The objects and letters were arranged in five rows, each containing 15 objects. The task was audio-recorded, and the time in seconds was calculated from the recording to be used in the analyses.

Three reading tests were included: word list reading using a standardized test of word list reading (Häyrynen et al., 1999), the number of correctly read words in 45 s was used as the score; non-word list reading based on Tests of Word Reading Efficiency (Torgesen et al., 1999), the number of correctly read non-words in 45 s was used as the score; pseudoword text reading (Eklund et al., 2015), number of correctly read words and total reading time were used as the scores. Writing to dictation was also assessed in which the participant heard 20 words and had to write them on a sheet of paper. The number of correctly written words was used as the score.

2.5. Data analysis

2.5.1. Behavioral analysis

The cumulative learning index for each audiovisual stimulus was

calculated based on the performance in the testing blocks. More specifically, for each auditory and visual components in the audiovisual stimuli, the learning index was defined as 0 if the participant pressed the wrong response button, and N (N = 1,2,3 ...) for the Nth time of successful learning (defined as correct response for both AVC and AVI for Learnable set, and correct response for AVX for Control set). The learning index for a specific audiovisual stimulus was the average learning index of the constituent auditory and visual components. The learning index was applied to the auditory only, visual only, and audiovisual stimuli in the same learning block prior to the testing block. For the incorrect button press after at least one successful learning, the learning index was defined as missing values (-1) and excluded from the analysis. These incorrect responses most likely reflect forgetting and lapses in attention to the task. The learning index gives a detailed quantification of the learning progress for each audiovisual stimulus; a learning index of N (greater than zero) indicates that successful learning happened on the Nth time when the audiovisual stimulus was presented.

The performance on the task (reaction time and accuracy) in the testing blocks for Learnable and Control stimuli were separately averaged by block (Fig. 2A). In total, there were 12 (Block Index 1–12) blocks on Day 1, and 6 (Block Index 13–18) blocks on Day 2. In addition, the reaction time was also examined using the cumulative learning index (see the previous paragraph) to show the detailed learning progress (Fig. 2B).

Based on the learning progress, indicated by the reaction time (see Fig. 2B), the participants acquire the letter-speech sound associations adequately after about 4 blocks of successful learning. The MEG data for Day 1 were therefore split over 3 learning stages (learning index = 0,1–4 and >4) for the audiovisual conditions in learning and testing conditions separately. For Day 2, the MEG data were averaged together, since the participants had already learned all the audiovisual pairs. For the different learning cues, we postulate that the participants were paying attention to them before learning and immediately following the first few successful learning trials. This could be related to the fact that the short-term memory of the audiovisual stimuli is not well stabilized and consolidated, which could be seen from the fast decrease of reaction time when learning index = 1–4 in Fig. 2B. Therefore similar learning-related cognitive processes were presented when learning index = 0–4 for the learning cue. To better capture the cognitive process in response to different learning cues, the MEG data were split into the following 3 parts for comparing the learning cues: learning index 0–4, learning index >4 on Day 1, and all the data on Day 2.

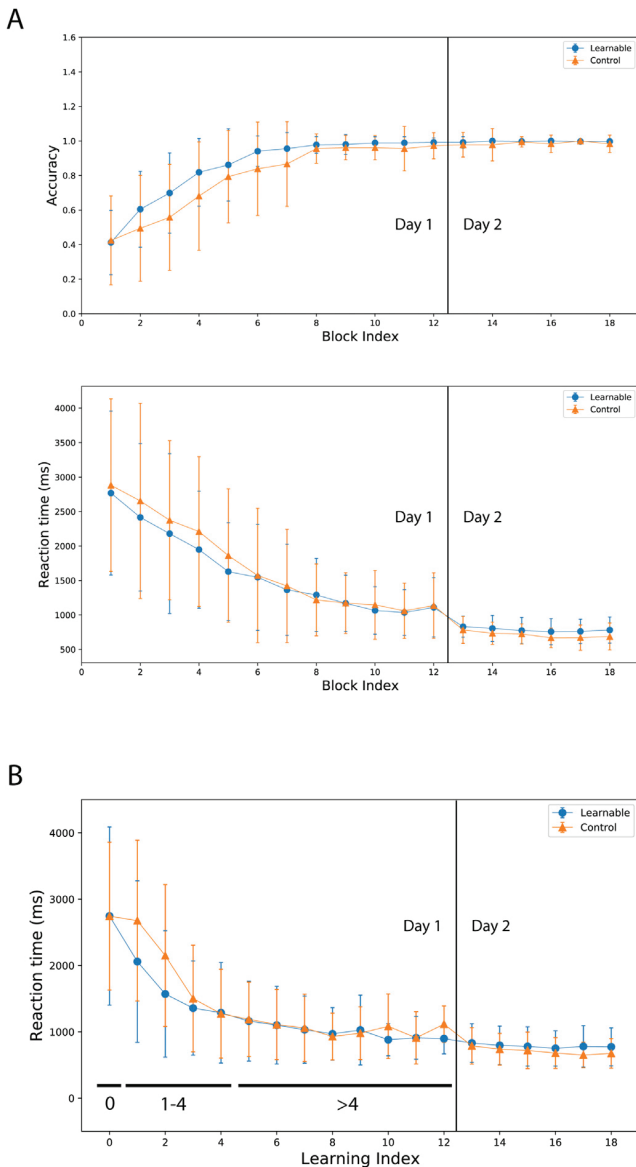


Fig. 2. Task performance curve (mean and standard deviation) for Learnable and Control stimuli during the two-day audiovisual learning experiment averaged by block (A) and learning index (B). A) Reaction time and accuracy averaged by block. In total, there are 12 blocks (Block Index 1–12) on Day 1 and 6 blocks (Block Index 13–18) on Day 2 with a vertical line in the figure separating the two days. B) Reaction time averaged by the cumulative learning index. A Learning Index of 0 indicates that the specific audiovisual pair has not been learned yet, and an integer number N greater than 0 indicates the N th time of successful learning in the testing block for the audiovisual pairs.

2.5.2. MEG data analysis

MEG data were first pre-processed with Maxfilter 3.0.17 to remove external noise interference and compensate for head movement during the recording, using the movement compensated temporal signal-space separation (tSSS) method (Taulu and Simola, 2006). Bad MEG channels were identified manually and were excluded; the time-series of the bad channels were reconstructed by Maxfilter.

Data were analyzed with MNE python (Gramfort et al., 2013) (version: 0.17.2). First, bad segments of MEG data were annotated and excluded from further analysis (Jas et al., 2018). MEG data were low-pass filtered at 40 Hz (zero-phase FIR filter design using the “hamming” window method). Fast ICA (Hyvärinen, 1999) was then used to remove

eye movement-related and cardiac artifacts. After applying ICA, data were segmented into epochs with 150 ms before and 1000 ms after stimulus onset. Bad epochs were first rejected based on peak-to-peak amplitude ($\text{grad} = 1500\text{e-}13 \text{ T/m}$, $\text{mag} = 5\text{e-}12 \text{ T}$) and then visually inspected in case of some remaining artifacts. Baseline correction was applied by subtracting the average response of the 150 ms prior to the stimulus onset from all data points throughout the epoch.

The multimodal interaction effects (i.e., processing of multi-modal stimuli is not merely the sum of auditory and visual stimuli presented separately) reflecting automatic and basic interaction between the auditory and visual processing were examined using the additive model (Audiovisual response = Auditory only response + Visual only response + Audiovisual interaction response). To calculate this regression analysis, the “linear_regression_raw” function in MNE Python was used. The interaction effects were calculated separately for the Learnable (LB) and Control (CT) stimuli by the linear regression analysis ($\text{AVC} = A_{\text{LB}} + V_{\text{LB}} + \text{Interaction}_{\text{LB}}$; $\text{AVX} = A_{\text{CT}} + V_{\text{CT}} + \text{Interaction}_{\text{CT}}$) for the three learning stages on Day 1 and Day 2. As can be seen from the model equations above, auditory response, visual response, and audiovisual interaction were the three predictors in the regression model. Since the MEG data was segmented into different learning stages, in order to get an adequate signal to noise ratio, a minimum number of 10 trials per average was used for including the MEG data in each regression analysis (Boudewyn et al., 2018; Luck, 2005).

The fsaverage brain template from Freesurfer (RRID: SCR_001847, Martinos Center for Biomedical Imaging, Charlestown, MA, United States) was used for source reconstruction since individual MRIs were not available in the present study. Coregistration was done between the digitized head points and the template brain with a 3-parameter scaling. The average of the three scaling parameters was calculated for the recordings on Day 1 and Day 2 and was used for coregistration for both days to ensure the use of the same forward head model for each participant across the two days.

Depth-weighted ($p = 0.8$) minimum-norm estimates (wMNE) (Hämäläinen and Ilmoniemi, 1994; Lin et al., 2006) were calculated for 10242 free-orientation sources per hemisphere. The dynamic statistical parametric maps (dSPM) (Dale et al., 2000) were used for noise normalization.

Region of interest analysis was used for comparing the estimated source activations to the three different audiovisual stimuli (AVC, AVI, AVX) in order to examine interaction effects in an analysis of variance (ANOVA) model. This was carried out for the training and testing blocks in 3 different learning stages on Day 1 and Day 2. Based on earlier literature (Karipidis et al., 2017; Xu et al., 2019; Rajj et al., 2000) brain dSPM source waveforms (500 ms–800 ms after stimulus onset) were extracted from left and right bank of the posterior superior temporal sulcus (labels: “bankssts”) (Calvert et al., 2001; Blomert, 2011; van Atteveldt et al., 2009; Xu et al., 2019; Beauchamp et al., 2004a; Wilson et al., 2018) as defined by the Desikan-Killiany Atlas (Desikan et al., 2006).

2.6. Statistical analyses

The audiovisual congruency effect was examined in a 3 (congruency: AVC, AVI, AVX) \times 2 (hemisphere: left, right) repeated-measures analysis of variance (ANOVA) for the estimated activity in the region of interest analysis. To further examine possible audiovisual learning effects related to unisensory processing and multisensory interaction, spatiotemporal cluster-based permutation tests (Maris and Oostenveld, 2007) were used for comparing Learnable and Control auditory, visual, and audiovisual interaction brain activations from the linear regression analysis based on the additive model.

Learning involves multiple cognitive components such as attention, active engagement, error feedback and memory formation and consolidation. In our experiment, these cognitive processes were reflected in the evoked responses elicited by the learning cues. To examine the cognitive

processes that lead to successful learning of the AV associations, brain responses to different learning cues (“YES”: ✓; “NO”: X; “UNKNOWN”: ☒) were also compared in pairs using the spatiotemporal cluster-based permutation tests. The number of permutations was set to 1000 for each test. The source data were downsampled to 200 Hz to reduce the computation time. The alpha threshold level was set to 0.05 for all tests.

Finally, to explore how much variance of the reading-related cognitive scores could be explained by the learning speed of Learnable and Control stimuli, correlation analysis (Pearson’s correlation coefficients) was carried out between the individual learning speed (average learning index of all Learnable and Control stimuli pairs in the twelfth block) on Day 1 and all the cognitive test scores (see the section above). The false discovery rate (FDR) was applied to correct the p-values in correlation analysis for the number of tests (Benjamini and Hochberg, 1995).

3. Results

3.1. Cognitive skills and experiment performance

Descriptive statistics of the participants’ cognitive skill measures are presented in Table 1.

As shown in Fig. 2 A, the participants were able to learn the correct associations between the auditory and visual stimuli and to differentiate the Control stimuli from the Learnable stimuli. This was indicated by an increase in accuracy scores and a decrease in reaction times over training blocks. A similar decrease and stabilization of reaction times can also be observed by examining the reaction time averaged by learning index for Learnable and Control stimuli (Fig. 2B). On Day 2, although the accuracy already reached ceiling level, there was a decrease of reaction time compared to the last block on Day 1 (Day 1 block 12: 1117 ms ± 436 ms vs. Day 2 block 13: 825 ms ± 153 ms).

3.2. Grand average

The grand average of both sensor and source-level brain activities for auditory, visual, and audiovisual conditions on Day 1 and Day 2 are shown in Fig. 3 (averaged across conditions). The activity patterns are typical for each stimulus type both in timing and in topography (localization). For example, the auditory responses were mainly localized around the perisylvian areas, and the visual responses were mainly localized in the occipital cortices.

Table 1

Descriptive statistics of the participants’ cognitive skill measures (N = 30). Standardized scores are shown within the parenthesis for block design, vocabulary, and digit span.

Cognitive tests	Mean	SD	Range
Block design	55.57 (12.1)	9.97 (3.97)	28 - 68 (2-19)
Vocabulary	39.47 (13.03)	7.49 (3.24)	23 - 49 (6-18)
Digit span	30.70 (12.43)	4.61 (2.67)	22 - 44 (7-19)
Nonword repetition, time (s)	11.30	2.44	6-16
NEPSY phonological processing	51.80	1.03	49-53
Rapid automatic naming(letters), time (s)	17.93	4.48	10-29
Rapid automatic naming(objects), time (s)	31.73	4.49	23-40
Word list reading, number of correct items	104.77	0.50	103-105
Word list reading, time (s)	70.27	10.83	53-98
Non-word list reading, number of correct items	71.07	6.31	58-86
Non-word text reading, time (s)	29.97	5.35	22-42
Non-word text reading, number of correct words	37.20	1.10	35-38

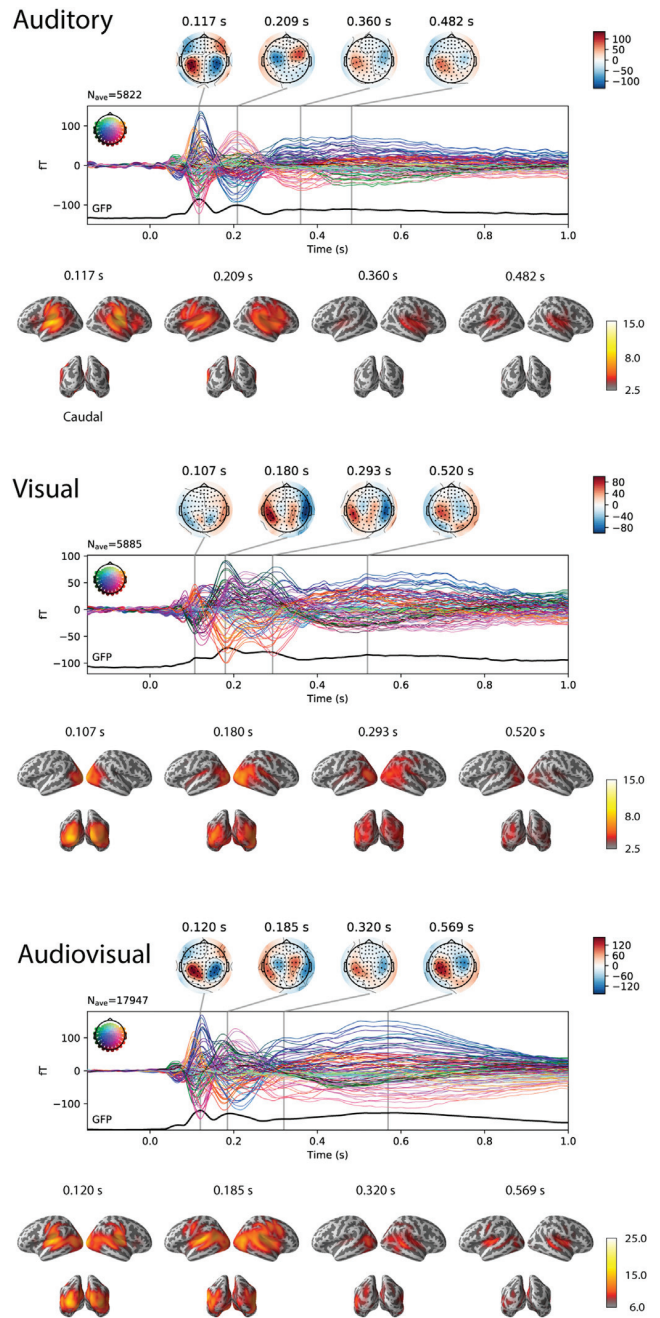


Fig. 3. Grand average of both sensor and source brain activities for auditory, visual, and audiovisual conditions on Day 1 and Day 2. Top of each panel: Grand average of event-related field waveforms and topographic plots of major peaks defined by global field power from magnetometer channels. Bottom of each panel: Brain source dSPM activation at the major peaks as defined in the sensor level topographic plots.

3.3. Congruency effects in the STC

Repeated-measures ANOVA revealed significant main effects of the congruency on Day 1 only after learning of letter-speech sound associations in the training blocks (learning index >4: $F(2, 52) = 4.81, p = 0.017$) and in the testing blocks (learning index 1-4: $F(2, 58) = 4.37, p = 0.022$; learning index >4: $F(2, 54) = 4.43, p = 0.022$), as well as on Day 2 ($F(2, 58) = 3.82, p = 0.034$) during the training blocks.

Post-hoc t-tests indicated that dSPM activation to the Control

audiovisual stimuli (AVX) was significantly lower ($p < 0.05$) than to the Learnable audiovisual stimuli (AVC and AVI) in the Day 1 training blocks (when learning index: >4) and to the audiovisual congruent stimuli (AVC) in the Day 2 training blocks. During the testing blocks when the learning index was 1–4, the incongruent audiovisual stimuli (AVI) elicited significantly higher ($p < 0.05$) activation than the Control audiovisual stimuli (AVX). The Learnable congruent audiovisual stimuli (AVC) elicited significantly higher ($p < 0.05$) activation than the Learnable incongruent audiovisual stimuli (AVI) and the Control stimuli (AVX) in the Day 1 testing blocks when learning index was greater than 4. In addition, there was a hemisphere main effect ($F(1, 29) = 7.48, p = 0.011$) with higher dSPM activation in the right hemisphere than the left hemisphere during the training blocks on Day 1 at the stage when the learning index was 1–4. The results of the congruency effect are summarized in Fig. 4.

3.4. Auditory responses (Learnable vs. control)

The learning effects were tested in the auditory modality by comparing the brain activations to the Learnable and Control auditory sounds from the time window of 100 ms–800 ms after the stimulus onset using spatiotemporal cluster-based permutation statistics. The statistical tests were carried out for the 3 different learning stages (Learning index

= 0, 1–4, >4 respectively) on Day 1 and the learned stage on Day 2. No significant differences were found for all the comparisons on Day 1 and Day 2.

3.5. Visual responses (Learnable vs. control)

The learning effects were tested in the visual modality by comparing the brain activations of the Learnable and Control visual letters from the time window of 100 ms–800 ms after the stimulus onset using spatiotemporal cluster-based permutation statistics. The statistical tests were carried between Learnable and Control stimuli for the three different learning stages (Learning index = 0, 1–4, >4 respectively) on Day 1 and the learned stage on Day 2. Significant differences were found between the Learnable and Control conditions on Day 1 when the learning index is greater than 4 ($p = 0.002, 455–795$ ms, left parietal, and occipital regions) and on Day 2 ($p = 0.001, 380–795$ ms, left parietal and occipital regions).

3.6. Audiovisual interaction effects (Learnable vs. control)

The learning effects were tested for the audiovisual interaction by comparing the audiovisual interaction brain activations of the Learnable (AVC- A_{LB} - V_{LB}) and Control (AVX- A_{CT} - V_{CT}) conditions from the time

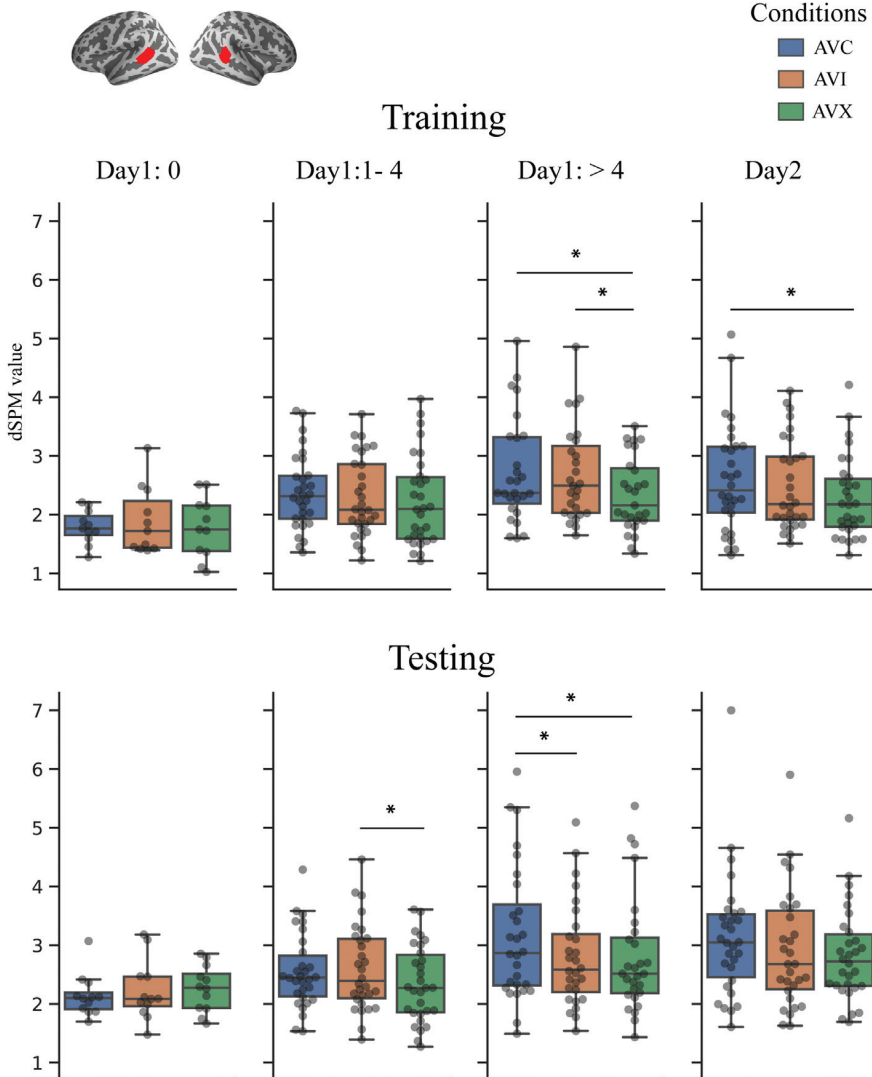


Fig. 4. Congruency effect in the region of interest (left and right posterior superior temporal sulcus) for Day 1 and Day 2 during the training and testing blocks. The data were divided into three stages on Day 1 based on the learning indexes: 0, which means the letter-speech sound association has not been learned, learning index 1–4, and learning index >4 . Since the participants had already learned all letter-speech sound pairs after Day 1, the data from all blocks on Day 2 were averaged. Results are shown in box plots with individual data marked as grey circles. Significant differences ($p < 0.05$) are marked by horizontal bars and asterisks (*).

window of 500 ms–800 ms after the stimulus onset using spatiotemporal cluster-based permutation statistics. The statistical comparisons between the Learnable and Control stimuli for the three different learning stages (Learning index = 0, 1–4, >4 respectively) on Day 1 and the learned stage on Day 2. There was a significant difference ($p = 0.019$, 500–680 ms, left parietal region) at the stage when the learning index was 1–4 on Day 1.

Results for the auditory, visual, and audiovisual interaction comparisons between Learnable and Control conditions are shown in Fig. 5.

3.7. Cortical responses to different learning cues

The cortical activities following the three different learning cues were compared in pairs using the spatiotemporal cluster-based permutation tests in the time window of 100 ms–800 ms for Day 1 (Learning index = 0–4, >4) and Day 2. There were significant differences between the 3 different learning cues when the learning index was between 0 and 4 on Day 1. Two clusters exceeding the threshold of randomization distribution under H_0 were found for the ✓ vs. X comparison, one ($p = 0.012$) in the left temporal regions in the time window of 300–490 ms and another ($p = 0.016$) in the right temporal regions in the time window of 295–550 ms. Two clusters were found for the ▨ vs. X comparison, one ($p = 0.008$) in the left temporal regions in the time window of 360–730 ms and another ($p = 0.036$) in the right temporal regions in the time window of 355–785 ms. Two clusters exceeding the randomization distribution under H_0 were found for the ✓ vs. ▨ comparison, one ($p = 0.040$) in the left temporal regions in the time window of 400–780 ms and another ($p = 0.037$) in the right temporal regions in the time window of 245–455 ms. In addition, there was a significant difference for the ▨ vs. X comparison ($p = 0.029$, 300–740 ms, left temporal, and occipital regions) when the learning index was greater than 4 on Day 1. No significant differences were found between the 3 different learning cues on Day 2. The results are shown in Fig. 6.

3.8. Correlations between cognitive skills and learning speed

Correlation analysis was carried out between learning speed (of the Learnable and Control stimuli) and cognitive test scores. After FDR correction, only the learning speed of the Control stimuli was significantly correlated with the time spent on RAN objects (FDR-corrected $p = 0.000168$). The results are shown in Fig. 7.

4. Discussion

This study investigated the grapheme-phoneme association learning in adults. The cortical dynamics during initial learning and memory consolidation after learning were captured in a two-day letter-speech sound learning experiment using MEG. In the experiment, two sets of audiovisual stimuli were used for training in which the letter-speech sound association could be learned in one set (Learnable), but not in the other set (Control), based on the different learning cues provided. The experiment was designed to dissociate the audiovisual processing and the grapheme-phoneme associative learning by consecutive presentations of, first, the audiovisual stimuli, and second, different learning cues. The participants' performance was monitored with trial-by-trial precision in the testing blocks after each learning block. This allowed us to examine the changes related to associative learning by comparing the Learnable and Control conditions at different learning stages. These comparisons revealed dynamic changes in the brain processes during multisensory learning and, most interestingly, during the processing of the learning cues.

Region of interest analysis was conducted for comparing the brain responses to the audiovisual stimuli in the Learnable (AVC, AVI) and Control (AVX) sets in the posterior superior temporal sulcus (pSTS) in the time window of 500–800 ms based on earlier studies (van Atteveldt et al., 2004, 2009; Beauchamp et al., 2004b; Raji et al., 2000; Karipidis et al., 2017; Xu et al., 2019; Wilson et al., 2018). As we expected, no significant differences were observed in the cortical responses to the different audiovisual stimuli before learning (learning index:0). The brain first started to process the audiovisual stimuli in the Learnable (AVI) and Control (AVX) sets differently in the early learning stages (learning index:1–4) in the testing blocks, and in the following training blocks (AVC > AVX and AVI > AVX) when learning index was >4 on Day 1. This suggested that the participants started to differentiate the Learnable and Control stimuli early in the learning process, which may reflect the easier differentiation of the two sets (Learnable and Control) compared to the learning of the audiovisual association within the Learnable set.

An effect of audiovisual congruency (AVC > AVI) in the left and right posterior superior temporal sulcus was found only at a later stage (learning index: >4) after the successful learning of letter-speech sound associations in the testing blocks on Day 1. This represents a brain level index of the learned associations. The stronger activation to congruent audiovisual stimuli than incongruent and control stimuli is in line with

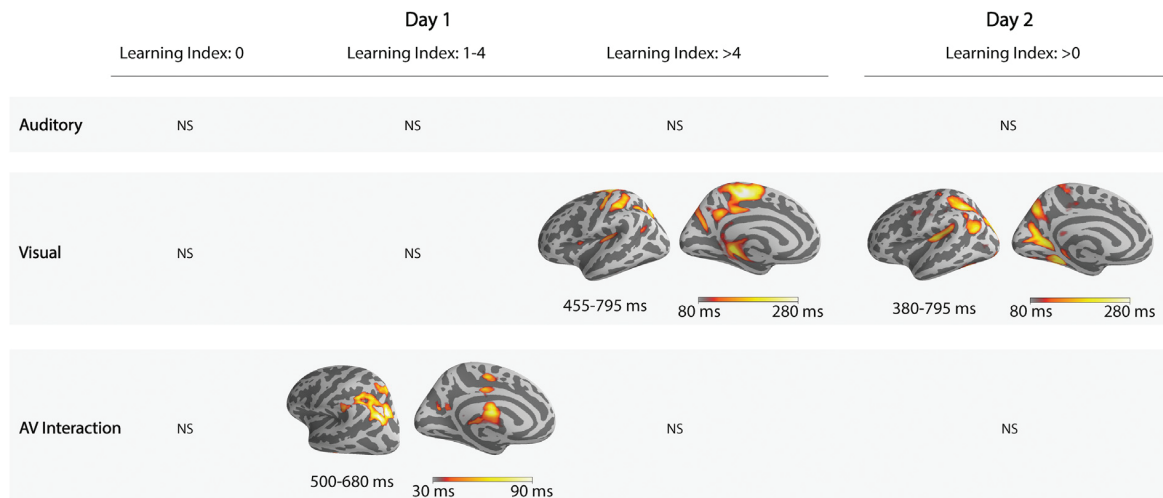


Fig. 5. Learning effect for the auditory, visual, and audiovisual interaction conditions between the Learnable and Control stimuli on Day 1 (learning index: 0, 1–4, >4) and on Day 2 using spatiotemporal cluster-based permutation tests at the source level. The cluster on the basis of which the null-hypothesis was rejected is represented on the cortical surface, and the temporal duration of the cluster is indicated underneath the cortical surface. The brightness of the color on the cortical surface is scaled by the temporal duration of the cluster. Non-significant results are marked with NS.

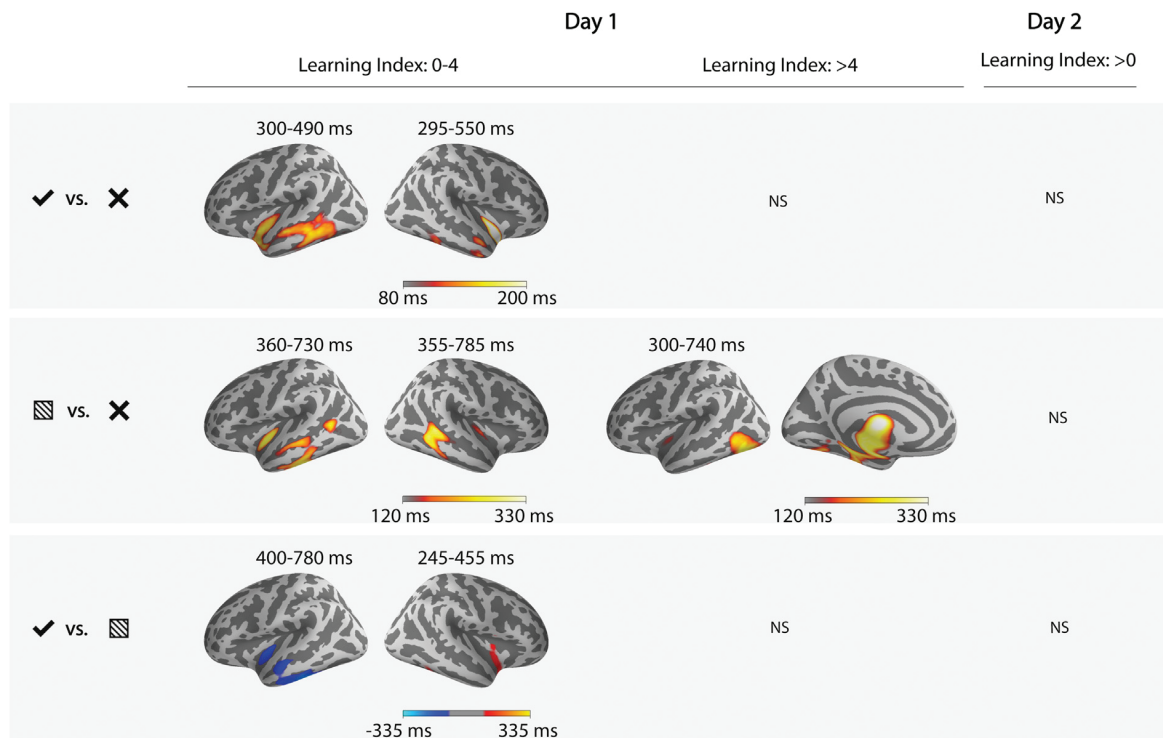


Fig. 6. Comparison of cortical activity following the different learning cues at different learning stages. The activities following the three different learning cues were compared in pairs using the spatiotemporal cluster-based permutation tests in the time window of 100 ms–800 ms. The MEG data were split into the following three parts: learning index 0–4, learning index >4 on Day 1, and all the data on Day 2. The time window of the cluster exceeding the threshold of randomization distribution under H0 is shown above each inflated brain. The colored region on the cortical surface is representing the cluster, and the brightness is scaled by its temporal duration. Warm color means the difference is greater than zero, and cold color means the difference is smaller than zero. Non-significant results are marked with NS.

congruency effects reported in earlier studies using similar letter-speech sound learning paradigm (Karipidis et al. 2017, 2018). In addition, a similar effect has also been consistently reported for over-learned letters in literate adults (van Atteveldt et al., 2004; Raj et al., 2000) and has been interpreted as a result of language-related audiovisual functional connections that have developed during learning to read (Van Atteveldt et al., 2004). However, this congruency effect in the testing blocks was absent on Day 2, possibly reflecting the effect of memory consolidation during sleep which could lead to the functional reorganization of multisensory memory (Rothschild, 2019). Such functional reorganization of cross-modal connections might still be incomplete only one day after the initial learning, but with more practice and repetition could lead to the automation of letter-speech sound integration in literate adults (Froyen et al., 2009). The congruency effect has also been shown to be dependent on the experimental task (e.g., active/passive or

explicit/implicit) (van Atteveldt et al., 2007; Blau et al., 2008). In our study, this is manifested as the different congruency effects for training and testing tasks at different learning stages in two days. For example, in contrast to the testing blocks, the brain responses to the Learnable audiovisual congruent (AVC) and Control (AVX) stimuli still showed a significant difference in the training blocks on Day 2. These differences most likely reflect the different cognitive processes during the training and testing blocks: e.g. active multisensory memory encoding was possibly mainly engaged during early training blocks, while memory retrieval and multisensory integration could only be possible after learning of audiovisual associations at later testing blocks. Therefore, these changes of congruency effects at different learning stages and during different tasks suggest dynamic characteristics of brain processes related to the newly-learned audiovisual associations.

Response to audiovisual congruence is not the only index reported in earlier literature that changes after learning the grapheme-phoneme associations. Therefore, more basic audiovisual interaction processes were also examined using the additive model (A + V vs. AV). The audiovisual interaction effect showed differences between the Learnable and Control conditions only at the early learning stage (learning index:1–4) on Day 1. Compared to the congruency effect, which is only possible after the crossmodal association has been learned (van Atteveldt et al., 2009), the (A + V vs. AV) comparison reflects more general form of cross-modal interaction, which has been shown to be important in children learning to read (Xu et al., 2018). The difference was maximal in the left parieto-occipital cortex, which has been indicated to be crucial for grapheme-phoneme mapping in learning to read (Sandak et al., 2004; Pugh et al., 2013; Bonte et al., 2017). Audiovisual interaction in the parieto-occipital cortex seems less often reported compared to the superior temporal cortex in many of the fMRI studies (Wilson et al., 2018; Van Atteveldt et al. 2004, 2009). Earlier MEG studies using similar contrasts based on additive models (A + V vs. AV) have shown

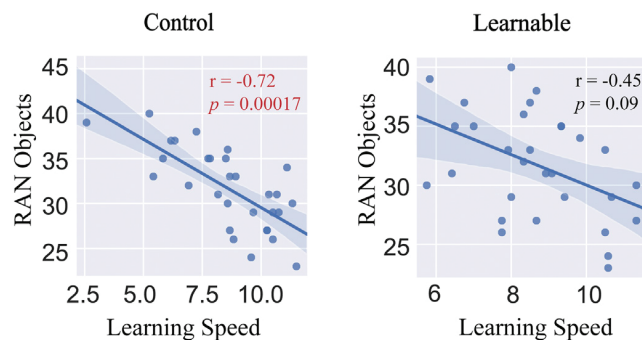


Fig. 7. Correlations between learning speed and RAN objects. Significant correlations were marked with red color text, non-significant correlations in black.

converging results in the parietal and occipital regions (Raij et al., 2000; Xu et al., 2018). Therefore this discrepancy could possibly be related to the different experimental paradigms and contrasts used in these studies to examine the audiovisual processes in learning to read. In the present study, the difference was only significant in the early learning stage, suggesting a transient learning process of actively combining auditory and visual information in the early learning stage for the Learnable set. In the later stages, after more stable multi-sensory memory representation was formed, no differences were found between the Learnable and Control stimuli.

Rather stable effects caused by learning of the audiovisual associations were found when examining the brain activation to the visual stimuli presented alone. The brain started to process the unimodally presented Learnable and Control letters differently at a later learning stage (learning index: >4), and this effect persisted on Day 2. The Learnable letters were closely linked to auditory stimuli (the phonemes) through repeated training on two days, whereas no such orthographic and phonological connection could be learned for the Control letters and phonemes. From the principle of Hebbian learning, seeing these Learnable letters alone should activate the phonemic representations of the letters. This seemed to occur after 4 repetitions of successful learning on Day 1 and continued to Day 2. Differences in phonological and orthographic processing of single letters vs. pseudo-letters have been reported (Bann and Herdman, 2016; Herdman and Takai, 2013) in both early (P1, N1, P2, and P3) and late (>300 ms) time windows using event-related potentials (ERPs). In our study, the time window of the cluster where responses to the Learnable and Control letters differed started from 455 ms on Day 1 and 380 ms on Day 2, which is relatively late compared to similar effects of learned letter vs. pseudo-letter comparisons in other studies (Herdman and Takai, 2013; Brem et al., 2018; Maurer et al., 2005). This might reflect the very early stages of learning captured by the present study, which might exhibit a slower processing speed of grapheme-phoneme mapping than the processing of well-established or over-learned associations. The spatial extent of the clusters for the Learnable and Control letter comparisons showed widespread distribution around the left temporoparietal, paracentral, and occipital regions. The temporoparietal (dorsal) circuits including the angular gyrus and supramarginal gyrus (SMG) in the inferior parietal lobule have been associated with grapheme to phoneme conversion (Pugh et al., 2000; Sandak et al., 2004; Taylor et al., 2014; Bonte et al., 2017) and are suggested to show brain changes in early reading acquisition (Dehaene et al., 2015; Pugh et al., 2001; Carreiras et al., 2015). Corroborating evidence comes from a training study in which participants were taught to read artificial words (componential learning of letter-sound associations) and name artificial objects over two days with differences observed in the occipitotemporal and parietal regions when reading an artificial orthography compared to naming artificial objects (Quinn et al., 2017; Naumer et al., 2009).

With regard to unimodally presented auditory stimuli, we did not find any differences in the brain responses to the Learnable and Control sets. This was not surprising because the auditory stimuli were Finnish phonemes, which are familiar for native Finnish participants and are already closely linked to the Finnish letters through years of experience. Therefore, each phoneme is mapped to at least one Finnish letter and a newly learned Georgian letter for the Learnable set. However, mapping of additional visual information to existing phonemic representations might not alter brain representations of the existing phonemes. Instead, associations between the phonemic representation and a new visual representation would be formed, which is what we saw for the processing of audiovisual stimuli.

Overall, the results related to brain responses of the audiovisual stimuli suggested that multisensory processing is very dynamic and depends on the different learning stages and tasks. On the other hand, the effects on brain responses to the unimodally presented letters seem to be more persistent after successful learning. These early dynamic processes have not been reported before since most earlier studies have examined

the multisensory or learning effects at one time point after training.

The findings described above are mostly brain-level indices of the learning results that occurred during the MEG recording. Of particular interest is, however, the brain mechanisms that lead to successful learning. Examining the brain responses to the three learning cues provided a unique window into brain processes that preceded the behavioral level of learning. Indeed, the brain responses to the learning cues were different mainly before and immediately after (learning index:0–4) behavioral learning could be observed on Day 1. On Day 2, after the audiovisual stimuli had been learned and consolidated, no differences were found between the brain activations to the three learning cues. In the training blocks, the auditory and visual information needs to be kept in working memory together with an initial weak association between the representations of the two different modalities after the presentation of audiovisual stimuli. The learning cue is processed first in the visual cortex (similar for all three cues) to extract the relevant information on the cross-modal relationship on the previously presented audiovisual stimuli, followed by the updating of that cross-modal relationship according to the cue. Both the associative learning process (reflected in the ✓ vs. X contrast) and non-associative learning process (reflected in the ☒ vs. X contrast) showed differences in the left and right middle and inferior temporal and some deeper brain sources (labeled as insula in the Desikan-Killiany atlas (Desikan et al., 2006) probably reflecting short-term memory encoding of multisensory information. Differences found in the deep brain sources near the bilateral medial temporal regions, as well as the insula, could reflect the working memory processes in the hippocampus and related areas (Yonelinas, 2013; Quak et al., 2015; Olson et al., 2006). Similar activation patterns have been reported in earlier studies; for example, the inferotemporal (IT) cortex has long been suggested to be important in forming associative long-term memory representations (Sakai and Miyashita, 1991; Miyashita and Hayashi, 2000) including audiovisual cross-modal associations (Gibson and Maunsell, 1997). In addition, the occipitotemporal junction and parahippocampal gyrus have been reported to show increased activation when learning arbitrary cross-modal associations (Tanabe et al., 2005; Skeide et al., 2018).

The difference between associative and non-associative learning (✓ vs. ☒) was mainly localized in parts of the left temporal and right insula regions. The decreased activation in left temporal regions might be related to the cross-modal memory encoding (Tanabe et al., 2005), and the increased activation in the right insula regions might be related to increased attention (Chen et al., 2015) for multisensory associative learning compared with non-associative learning. The relatively late time window of the effect (after about 300 ms) also suggests a multisensory working memory process engaged after the basic sensory processing of the learning cues. Therefore, cross-modal associative and non-associative learning utilized largely overlapping left and right middle and inferior temporal and deep brain regions (possibly reflecting activity from or activity modulated by, for example, insula or hippocampus) when using the 'X' cue as the baseline. However, these two types of learning also showed differential activation strength in both hemispheres, which probably reflects the unique cognitive processes in associative learning.

For the correlation analysis between learning speed and cognitive skills, we found that audiovisual non-associative learning speed was significantly correlated with the rapid naming of objects. Rapid naming ability, which is a robust behavioral precursor of reading fluency across various languages (Moll et al., 2014; Kirby et al., 2010), seems to be important for fast non-associative learning. This result is consistent with other studies using artificial language learning studies (Karipidis et al., 2017, 2018; Aravena et al., 2018), although this one correlation surviving the multiple comparison correction should be regarded with caution as the main variable of interest (the learning speed for the Learnable items) did not show significant correlations to the naming speed.

Based on findings from the present and previous studies, we formulate a global sketch regarding the learning of letter-speech sound

associations in Fig. 8. In this figure, the auditory (e.g., the sound of /a/) and visual (e.g., letter a) sensory inputs are first processed in the primary auditory and visual cortices. The auditory features of the stimuli are then combined to form more abstract representations most likely in the superior temporal regions in both early and late time window as reflected for example by the auditory P2 response (Hämäläinen et al., 2019) and late sustained responses (Ceponiene et al., 2005). The visual features have been combined and formed an abstract representation when the visual information is processed along the vOT cortex known to respond to orthographic regularities. The auditory and visual information are then integrated in the multisensory areas in the superior temporal cortex (van Atteveldt et al., 2009; Raji et al., 2000; Beauchamp et al., 2004b) to form a coherent audiovisual object at a relatively late time window after the auditory and visual inputs are processed (see (van Atteveldt et al., 2009) for a functional neuro-anatomical model of letter-speech sound integration in literate adults).

During the initial learning stage, the audiovisual representation is encoded, and short-term memory of the audiovisual objects are stored in

the middle and inferior temporal and possibly also in the medial temporal (e.g., hippocampus) regions (Quinn et al., 2017). Frontal regions have been suggested to be involved in many aspects of language-related processes including those related to syntax and semantics (Skeide et al., 2014; Vigliocco, 2000), as well as the top-down control mechanism during language learning (Mei et al., 2014; Skeide and Friederici, 2016). For example, the frontal regions are involved in the selection of cross-modal features (Hämäläinen et al., 2019; Calvert et al., 2001) to combine and direct attention to the relevant learning cues. In addition, parietal regions also receive visual inputs (of letters) from the occipital regions and might be involved in storing the corresponding phonological representation of the letters by interacting with the multisensory superior temporal cortex during the early stages of learning. As learning progresses, changes have been reported to occur in vOT (Quinn et al., 2017; Madec et al., 2016; Hashimoto and Sakai, 2004; Brem et al. 2010, 2018) and dorsal pathway (Taylor et al. 2014, 2017; Hashimoto and Sakai, 2004; Mei et al. 2014, 2015) as well as the STC (Hämäläinen et al., 2019; Karipidis et al. 2017, 2018; Madec et al., 2016) for forming optimal

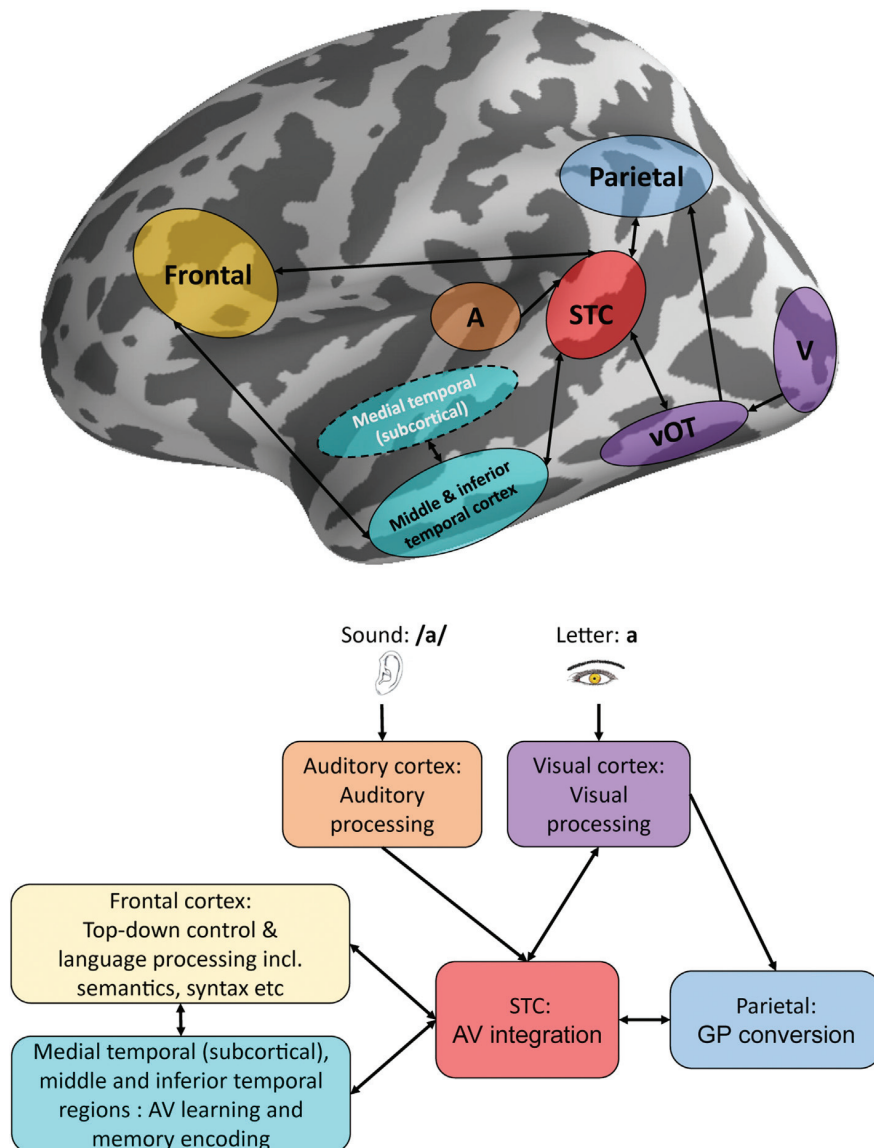


Fig. 8. Schematic diagram of the possible network involved in the learning of letters-speech sound associations. A = Auditory cortex, V = Visual cortex, STC = superior temporal cortex, vOT = ventral occipitotemporal cortex, GP = Grapheme-phoneme.

cortical representation and automatic processing of the audiovisual objects.

There are several limitations in our study. First, deep brain structures such as the medial temporal system (including the hippocampus) play a crucial role in the learning and memory processes as reported by numerous studies (Jarrard, 1993; Brasted et al., 2003; Mayes et al., 2007; Axmacher et al., 2008). One recent fMRI study has also demonstrated that reading-related reorganization could occur at the level of the brainstem and the thalamus (Skeide et al., 2017). MEG might not be optimal to localize the brain activity within these deep brain regions due to the decreased signal-to-noise ratio (SNR) as a function of source depth. However, evidence suggests that hippocampal activities could be captured with MEG (Ruzich et al., 2019; Attal and Schwartz, 2013), especially during learning and memory tasks (Taylor et al., 2012; Backus et al., 2016; Shah-Basak et al., 2018). In our study, there seemed to be some activity related to the processing of the learning cues from the deep brain sources. However, due to the limited SNR and spatial resolution in MEG, caution should be taken when interpreting these results, and particularly the localization regarding the medial temporal sources. Another limitation of our study relates to the lack of individual structural magnetic resonance images, which could potentially lead to poorer source localization accuracy and false-positive activation (Supek and Aine, 2014).

The current study tracked the learning process in two days; ideally, it would be interesting to track the learning process over a longer period (e.g., one week). In this study, our greatest interest was to investigate the initial stages of learning using a paradigm simulating the situation when children typically learn letters instead of tracking long-term changes in brain activity. Findings from the present study on brain dynamics during letter-speech sound learning could provide new information on potential mechanisms leading to long-term learning outcomes. It is very likely that similar brain networks are recruited for learning letter-speech sound associations in both children and adults since this process utilizes a more general audiovisual object association learning mechanism which is also essential in everyday life. It would be interesting to further investigate the potential difference in behavioral (e.g. learning speed) and brain level differences for adults and children. Learning also involves interaction and communication between different brain regions. Thus brain connectivity would be an interesting approach. The current study has identified important brain regions (hubs) and time windows that could be useful for future studies that use optimal experimental design for connectivity analysis on learning.

In conclusion, in this study, we have successfully captured some of the brain dynamics of learning grapheme-phoneme associations using a well-controlled audiovisual paradigm. Audiovisual processing showed fast and dynamic changes across different learning stages over two days and was modulated by the effect of overnight memory consolidation during sleep. Newly-learned letters that were associated with specific phonemes showed stronger activation along the dorsal pathway, probably reflecting the grapheme to the phoneme conversion process. We also identified other neural processes, for example, in the middle and inferior temporal cortices, that are important for multisensory learning and cross-modal memory encoding. Letter-speech sound learning deficit has been reported as a key factor for dyslexia in studies using artificial letter training paradigms (Aravena et al. 2013, 2018; Karipidis et al. 2017, 2018). Findings from the present study could provide a better understanding of neural dynamics that underpin grapheme-phoneme learning and could be used to find specific bottlenecks in learning cross-modal associations.

Data and Code Availability Statement

The associated MEG dataset has been shared in the Openneuro platform (<https://openneuro.org/datasets/ds002598>).

The analysis scripts are available in Github repository (<https://github.com/weiyongxu/Grapheme-phoneme-Learning.git>).

CRediT authorship contribution statement

Weiyong Xu: Conceptualization, Methodology, Software, Investigation, Data curation, Formal analysis, Visualization, Writing - original draft. **Orsolya Beatrix Kolozsvári:** Writing - original draft, Writing - review & editing. **Robert Oostenfeld:** Methodology, Writing - original draft, Supervision. **Jarmo Arvid Hämäläinen:** Conceptualization, Writing - original draft, Writing - review & editing, Project administration, Supervision.

Acknowledgements

We would like to thank Suvi Karjalainen, Aino Sorsa and Ainomaija Laitinen for their help in data collection. This work has been supported by the European Union projects ChildBrain (Marie Curie Innovative Training Networks, no. 641652), Predictable (Marie Curie Innovative Training Networks, no. 641858) and the Academy of Finland (MultiLeTe #292466). The authors declare no conflicts of interest.

References

- Aravena, S., Snellings, P., Tijms, J., van der Molen, M.W., 2013. A lab-controlled simulation of a letter-speech sound binding deficit in dyslexia. *J. Exp. Child Psychol.* 115 (4), 691–707.
- Aravena, S., Tijms, J., Snellings, P., van der Molen, M.W., 2018. Predicting individual differences in reading and spelling skill with artificial script-based letter-speech sound training. *J. Learn. Disabil.* 51 (6), 552–564.
- Attal, Y., Schwartz, D., 2013. Assessment of subcortical source localization using deep brain activity imaging model with minimum norm operators: a MEG study. *PLoS One* 8 (3).
- Axmacher, N., Elger, C.E., Fell, J., 2008. Ripples in the medial temporal lobe are relevant for human memory consolidation. *Brain: J. Neurol.* 131 (Pt 7), 1806–1817.
- Backus, A.R., Schoffelen, J.-M., Szebényi, S., Hanslmayr, S., Doeller, C.F., 2016. Hippocampal-prefrontal theta oscillations support memory integration. *Curr. Biol.: CB* 26 (4), 450–457.
- Bann, S.A., Herdman, A.T., 2016. Event related potentials reveal early phonological and orthographic processing of single letters in letter-detection and letter-rhyme paradigms. *Front. Hum. Neurosci.* 10, 176.
- Beauchamp, M.S., Argall, B.D., Bodurka, J., Duyn, J.H., Martin, A., 2004a. Unraveling multisensory integration: patchy organization within human STS multisensory cortex. *Nat. Neurosci.* 7 (11), 1190–1192.
- Beauchamp, M.S., Lee, K.E., Argall, B.D., Martin, A., 2004b. Integration of auditory and visual information about objects in superior temporal sulcus. *Neuron* 41 (5), 809–823.
- Benjamini, Y., Hochberg, Y., 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. Roy. Stat. Soc. B Stat. Methodol.* 57 (1), 289–300.
- Ben-Shachar, M., Dougherty, R.F., Deutsch, G.K., Wandell, B.A., 2011. The development of cortical sensitivity to visual word forms. *J. Cognit. Neurosci.* 23 (9), 2387–2399.
- Blau, V., Reithler, J., van Atteveldt, N., Seitz, J., Gerretsen, P., Goebel, R., Blomert, L., 2010. Deviant processing of letters and speech sounds as proximate cause of reading failure: a functional magnetic resonance imaging study of dyslexic children. *Brain: J. Neurol.* 133 (Pt 3), 868–879.
- Blau, V., Van Atteveldt, N., Formisano, E., Goebel, R., Blomert, L., 2008. Task-irrelevant visual letters interact with the processing of speech sounds in heteromodal and unimodal cortex. *Eur. J. Neurosci.* 28 (3), 500–509.
- Blomert, L., 2011. The neural signature of orthographic-phonological binding in successful and failing reading development. *Neuroimage* 57 (3), 695–703.
- Bönstrup, M., Iturrate, I., Thompson, R., Cruciani, G., Censor, N., Cohen, L.G., 2019. A rapid form of offline consolidation in skill learning. *Curr. Biol.: CB* 29 (8), 1346–1351 e4.
- Bonte, M., Correia, J.M., Keetels, M., Vroomen, J., Formisano, E., 2017. Reading-induced shifts of perceptual speech representations in auditory cortex. *Sci. Rep.* 7 (1), 5143.
- 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), e13049.
- Brasted, P.J., Bussey, T.J., Murray, E.A., Wise, S.P., 2003. Role of the hippocampal system in associative learning beyond the spatial domain. *Brain: J. Neurol.* 126 (Pt 5), 1202–1223.
- Brem, S., Bach, S., Kucian, K., Kujala, J.V., Guttorm, T.K., Martin, E., Lyytinen, H., Brandeis, D., Richardson, U., 2010. Brain sensitivity to print emerges when children learn letter-speech sound correspondences. *Proc. Natl. Acad. Sci. U. S. A.* 107 (17), 7939–7944.
- Brem, S., Hunkeler, E., Mächler, M., Kronschnabel, J., Karipidis, I.L., Pleisch, G., Brandeis, D., 2018. Increasing expertise to a novel script modulates the visual N1 ERP in healthy adults. *IJBD (Int. J. Behav. Dev.)* 42 (3), 333–341.
- Calvert, G.A., 2001. Crossmodal processing in the human brain: insights from functional neuroimaging studies. *Cerebr. Cortex* 11 (12), 1110–1123.
- Calvert, G.A., Hansen, P.C., Iversen, S.D., Brammer, M.J., 2001. Detection of audio-visual integration sites in humans by application of electrophysiological criteria to the BOLD effect. *Neuroimage* 14 (2), 427–438.

- Carreiras, M., Quiñones, I., Hernández-Cabrera, J.A., Duñabeitia, J.A., 2015. Orthographic coding: brain activation for letters, symbols, and digits. *Cerebr. Cortex* 25 (12), 4748–4760.
- Centanni, T.M., Norton, E.S., Park, A., Beach, S.D., Halverson, K., Ozernov-Palchik, O., Gaab, N., Gabrieli, J.D.E., 2018. Early development of letter specialization in left fusiform is associated with better word reading and smaller fusiform face area. *Dev. Sci.* 21 (5), e12658.
- Ceponiene, R., Alku, P., Westerfield, M., Torki, M., Townsend, J., 2005. ERPs differentiate syllable and nonphonetic sound processing in children and adults. *Psychophysiology* 42 (4), 391–406.
- Chen, T., Michels, L., Supekar, K., Kochalka, J., Ryali, S., Menon, V., 2015. Role of the anterior insular cortex in integrative causal signaling during multisensory auditory-visual attention. *Eur. J. Neurosci.* 41 (2), 264–274.
- 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.
- Davis, M.H., Di Betta, A.M., Macdonald, M.J.E., Gaskell, M.G., 2009. Learning and consolidation of novel spoken words. *J. Cognit. Neurosci.* 21 (4), 803–820.
- Dehaene, S., Cohen, L., 2011. The unique role of the visual word form area in reading. *Trends Cognit. Sci.* 15 (6), 254–262.
- Dehaene, S., Cohen, L., Morais, J., Kolinsky, R., 2015. Illiterate to literate: behavioural and cerebral changes induced by reading acquisition. *Nat. Rev. Neurosci.* 16 (4), 234–244.
- Dehaene, S., Cohen, L., Sigman, M., Vinckier, F., 2005. The neural code for written words: a proposal. *Trends Cognit. Sci.* 9 (7), 335–341.
- Dehaene, S., Pegado, F., Braga, L.W., Ventura, P., Nunes Filho, G., Jobert, A., Dehaene-Lambertz, G., Kolinsky, R., Morais, J., Cohen, L., 2010. How learning to read changes the cortical networks for vision and language. *Science* 330 (6009), 1359–1364.
- Denckla, M.B., Rudel, R.G., 1976. Rapid “automatized” naming (RAN): dyslexia differentiated from other learning disabilities. *Neuropsychologia* 14 (4), 471–479.
- Desikan, R.S., Ségonne, F., Fischl, B., Quinn, B.T., Dickerson, B.C., Blacker, D., Buckner, R.L., Dale, A.M., Maguire, R.P., Hyman, B.T., Albert, M.S., Killiany, R.J., 2006. An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *Neuroimage* 31 (3), 968–980.
- Diekemann, S., Born, J., 2010. The memory function of sleep. *Nat. Rev. Neurosci.* 11 (2), 114–126.
- Dudai, Y., 2012. The restless engram: consolidations never end. *Annu. Rev. Neurosci.* 35, 227–247.
- Eklund, K., Torppa, M., Aro, M., Leppänen, P.H.T., Lyytinen, H., 2015. Literacy skill development of children with familial risk for dyslexia through grades 2, 3, and 8. *J. Educ. Psychol.* 107 (1), 126–140.
- Froyen, D.J.W., Bonte, M.L., van Atteveldt, N., Blomert, L., 2009. The long road to automation: neurocognitive development of letter-speech sound processing. *J. Cognit. Neurosci.* 21 (3), 567–580.
- Froyen, D., Van Atteveldt, N., Bonte, M., Blomert, L., 2008. Cross-modal enhancement of the MMN to speech-sounds indicates early and automatic integration of letters and speech-sounds. *Neurosci. Lett.* 430 (1), 23–28.
- Froyen, D., Willems, G., Blomert, L., 2011. Evidence for a specific cross-modal association deficit in dyslexia: an electrophysiological study of letter-speech sound processing. *Dev. Sci.* 14 (4), 635–648.
- Gibson, J.R., Maunsell, J.H., 1997. Sensory modality specificity of neural activity related to memory in visual cortex. *J. Neurophysiol.* 78 (3), 1263–1275.
- Gramfort, A., Luessi, M., Larson, E., Engemann, D.A., Strohmeier, D., Brodbeck, C., Goj, R., Jas, M., Brooks, T., Parkkonen, L., Hämäläinen, M., 2013. MEG and EEG data analysis with MNE-Python. *Front. Neurosci.* 7, 267.
- Hämäläinen, J.A., Parviainen, T., Hsu, Y.-F., Salmelin, R., 2019. Dynamics of brain activation during learning of syllable-symbol paired associations. *Neuropsychologia* 129, 93–103.
- Hämäläinen, M.S., Ilmoniemi, R.J., 1994. Interpreting magnetic fields of the brain: minimum norm estimates. *Med. Biol. Eng. Comput.* 32 (1), 35–42.
- Hashimoto, R., Sakai, K.L., 2004. Learning letters in adulthood: direct visualization of cortical plasticity for forming a new link between orthography and phonology. *Neuron* 42 (2), 311–322.
- Häyrynen, T., Serenius-Sirve, S., Korkman, M., 1999. Lukilasse. Lukemisen, Kirjoittamisen Ja Laskemisen Seulontatutkimus Peruskoulun Ala-Asteen Luokille, pp. 1–6.
- Hebschs, M., Wing, E., Ryan, J., Gilboa, A., 2019. Rapid cortical plasticity supports long-term memory formation. *Trends Cognit. Sci.* 23 (12), 989–1002.
- Herdman, A.T., Takai, O., 2013. Paying attention to orthography: a visual evoked potential study. *Front. Hum. Neurosci.* 7, 199.
- Holloway, I.D., van Atteveldt, N., Blomert, L., Ansari, D., 2015. Orthographic dependency in the neural correlates of reading: evidence from audiovisual integration in English readers. *Cerebr. Cortex* 25 (6), 1544–1553.
- Hyvärinen, A., 1999. Fast and robust fixed-point algorithms for independent component analysis. *IEEE Trans. Neural Network.* 10 (3), 626–634.
- Jarrard, L.E., 1993. On the role of the hippocampus in learning and memory in the rat. *Behav. Neural Biol.* 60 (1), 9–26.
- Jas, M., Larson, E., Engemann, D.A., Leppäkangas, J., Taulu, S., Hämäläinen, M., Gramfort, A., 2018. A reproducible MEG/EEG group study with the MNE software: recommendations, quality assessments, and good practices. *Front. Neurosci.* 12, 530.
- Karipidis, I., Pleisch, G., Brandeis, D., Roth, A., Röthlisberger, M., Schneebeli, M., Walitza, S., Brem, S., 2018. Simulating reading acquisition: the link between reading outcome and multimodal brain signatures of letter-speech sound learning in prereaders. *Sci. Rep.* 8 (1), 7121.
- Karipidis, I., Pleisch, G., Röthlisberger, M., Hofstetter, C., Dornbierer, D., Stämpfli, P., Brem, S., 2017. Neural initialization of audiovisual integration in prereaders at varying risk for developmental dyslexia. *Hum. Brain Mapp.* 38 (2), 1038–1055.
- Kirby, J.R., Georgiou, G.K., Martinussen, R., Parrila, R., Bowers, P., Landerl, K., 2010. Naming speed and reading: from prediction to instruction. *Read. Res. Q.* 45 (3), 341–362.
- Klinzing, J.G., Niethard, N., Born, J., 2019. Mechanisms of systems memory consolidation during sleep. *Nat. Neurosci.* 22 (10), 1598–1610.
- Korkman, M., Kirk, U., Kemp, S., 1998. A Developmental NeuroPsychological Assessment (NEPSY), vol. 64. Psychological Corporation, San Antonio, TX, pp. 156–157.
- Korkman, M., Kirk, U., Kemp, S., 2007. NEPSY—Second Edition (NEPSY-II). Harcourt Assessment, San Antonio, TX.
- Lerma-Usabiaga, G., Carreiras, M., Paz-Alonso, P.M., 2018. Converging evidence for functional and structural segregation within the left ventral occipitotemporal cortex in reading. *Proc. Natl. Acad. Sci. Unit. States Am.* 115 (42), E9981–E9990.
- Lieberman, A.M., 1992. The relation of speech to reading and writing. *Adv. Psychol. Amsterdam* 94, 167–167.
- Lieberman, P., 2006. Toward an Evolutionary Biology of Language. Harvard University Press.
- Lin, F.-H., Witzel, T., Ahlfors, S.P., Stufflebeam, S.M., Belliveau, J.W., Hämäläinen, M.S., 2006. Assessing and improving the spatial accuracy in MEG source localization by depth-weighted minimum-norm estimates. *Neuroimage* 31 (1), 160–171.
- Luck, S.J., 2005. An Introduction to Event Related Potentials and Their Neural Origins. An Introduction to the Event Related Potential Technique, p. 11.
- Madeç, S., Le Goff, K., Anton, J.-L., Longcamp, M., Velay, J.-L., Nazarian, B., Roth, M., Courrieu, P., Grainger, J., Rey, A., 2016. Brain correlates of phonological recoding of visual symbols. *Neuroimage* 132, 359–372.
- Maris, E., Oostenveld, R., 2007. Nonparametric statistical testing of EEG- and MEG-data. *J. Neurosci. Methods* 164 (1), 177–190.
- Maurer, U., Brem, S., Bucher, K., Brandeis, D., 2005. Emerging neurophysiological specialization for letter strings. *J. Cognit. Neurosci.* 17 (10), 1532–1552.
- Mayes, A., Montaldi, D., Migo, E., 2007. Associative memory and the medial temporal lobes. *Trends Cognit. Sci.* 11 (3), 126–135.
- McClelland, J.L., McNaughton, B.L., O’Reilly, R.C., 1995. Why there are complementary learning systems in the hippocampus and neocortex: insights from the successes and failures of connectionist models of learning and memory. *Psychol. Rev.* 102 (3), 419–457.
- Mei, L., Xue, G., Lu, Z.-L., He, Q., Wei, M., Zhang, M., Dong, Q., Chen, C., 2015. Native language experience shapes neural basis of addressed and assembled phonologies. *Neuroimage* 114, 38–48.
- Mei, L., Xue, G., Lu, Z.L., He, Q., Zhang, M., Wei, M., et al., 2014. Artificial language training reveals the neural substrates underlying addressed and assembled phonologies. *PLoS One* 9 (3).
- Miyashita, Y., Hayashi, T., 2000. Neural representation of visual objects: encoding and top-down activation. *Curr. Opin. Neurobiol.* 10 (2), 187–194.
- Moll, K., Ramus, F., Bartling, J., Bruder, J., Kunze, S., Neuhoff, N., et al., 2014. Cognitive mechanisms underlying reading and spelling development in five European orthographies. *Learn. Instruct.* 29, 65–77.
- Naumer, M.J., Doehrmann, O., Müller, N.G., Muckli, L., Kaiser, J., Hein, G., 2009. Cortical plasticity of audio-visual object representations. *Cerebr. Cortex* 19 (7), 1641–1653.
- Olson, I.R., Moore, K.S., Stark, M., Chatterjee, A., 2006. Visual working memory is impaired when the medial temporal lobe is damaged. *J. Cognit. Neurosci.* 18 (7), 1087–1097.
- Pleisch, G., Karipidis, I.L., Brauchli, C., Röthlisberger, M., Hofstetter, C., Stämpfli, P., Walitza, S., Brem, S., 2019. Emerging neural specialization of the ventral occipitotemporal cortex to characters through phonological association learning in preschool children. *Neuroimage* 189, 813–831.
- Price, C.J., Devlin, J.T., 2011. The interactive account of ventral occipitotemporal contributions to reading. *Trends Cognit. Sci.* 15 (6), 246–253.
- Pugh, K.R., Landi, N., Preston, J.L., Mencl, W.E., Austin, A.C., Sibley, D., Fulbright, R.K., Seidenberg, M.S., Grigorenko, E.L., Constable, R.T., Molfese, P., Frost, S.J., 2013. The relationship between phonological and auditory processing and brain organization in beginning readers. *Brain Lang.* 125 (2), 173–183.
- Pugh, K.R., Mencl, W.E., Jenner, A.R., Katz, L., Frost, S.J., Lee, J.R., Shaywitz, S.E., Shaywitz, B.A., 2000. Functional neuroimaging studies of reading and reading disability (developmental dyslexia). *Ment. Retard. Dev. Disabil. Res. Rev.* 6 (3), 207–213.
- Pugh, K.R., Mencl, W.E., Jenner, A.R., Katz, L., Frost, S.J., Lee, J.R., Shaywitz, S.E., Shaywitz, B.A., 2001. Neurobiological studies of reading and reading disability. *J. Commun. Disord.* 34 (6), 479–492.
- Quak, M., London, R.E., Talsma, D., 2015. A multisensory perspective of working memory. *Front. Hum. Neurosci.* 9, 197.
- Quinn, C., Taylor, J.S.H., Davis, M.H., 2017. Learning and retrieving holistic and componential visual-verbal associations in reading and object naming. *Neuropsychologia* 98, 68–84.
- Raij, T., Uutela, K., Hari, R., 2000. Audiovisual integration of letters in the human brain. *Neuron* 28 (2), 617–625.
- Richlan, F., 2019. The functional neuroanatomy of letter-speech sound integration and its relation to brain abnormalities in developmental dyslexia. *Front. Hum. Neurosci.* 13, 21.
- Rothschild, G., 2019. The transformation of multi-sensory experiences into memories during sleep. *Neurobiol. Learn. Mem.* 160, 58–66.
- Ruzich, E., Crespo-García, M., Dalal, S.S., Schneiderman, J.F., 2019. Characterizing hippocampal dynamics with MEG: a systematic review and evidence-based guidelines. *Hum. Brain Mapp.* 40 (4), 1353–1375.

- Sakai, K., Miyashita, Y., 1991. Neural organization for the long-term memory of paired associates. *Nature* 354 (6349), 152–155.
- Sandak, R., Mencl, W.E., Frost, S.J., Pugh, K.R., 2004. The neurobiological basis of skilled and impaired reading: recent findings and new directions. *Sci. Stud. Read.* 8 (3), 273–292.
- Shah-Basak, P.P., Urbain, C., Wong, S., da Costa, L., Pang, E.W., Dunkley, B.T., Taylor, M.J., 2018. Concussion alters the functional brain processes of visual attention and working memory. *J. Neurotrauma* 35 (2), 267–277.
- Skeide, M.A., Brauer, J., Friederici, A.D., 2014. Syntax gradually segregates from semantics in the developing brain. *Neuroimage* 100, 106–111.
- Skeide, M.A., Evans, T.M., Mei, E.Z., Abrams, D.A., Menon, V., 2018. Neural signatures of co-occurring reading and mathematical difficulties. *Dev. Sci.* 21 (6), e12680.
- Skeide, M.A., Friederici, A.D., 2016. The ontogeny of the cortical language network. *Nat. Rev. Neurosci.* 17 (5), 323–332.
- Skeide, M.A., Kumar, U., Mishra, R.K., Tripathi, V.N., Guleria, A., Singh, J.P., et al., 2017. Learning to read alters cortico-subcortical cross-talk in the visual system of illiterates. *Sci. Adv.* 3 (5), e1602612.
- Stein, B.E., Stanford, T.R., 2008. Multisensory integration: current issues from the perspective of the single neuron. *Nat. Rev. Neurosci.* 9 (4), 255–266.
- Supek, S., Aine, C.J., 2014. Magnetoencephalography: from Signals to Dynamic Cortical Networks. Springer.
- Tanabe, H.C., Honda, M., Sadato, N., 2005. Functionally segregated neural substrates for arbitrary audiovisual paired-association learning. *J. Neurosci.: Off. J. Soc. Neurosci.* 25 (27), 6409–6418.
- Taulu, S., Simola, J., 2006. Spatiotemporal signal space separation method for rejecting nearby interference in MEG measurements. *Phys. Med. Biol.* 51 (7), 1759–1768.
- Taylor, J.S.H., Davis, M.H., Rastle, K., 2017. Comparing and validating methods of reading instruction using behavioural and neural findings in an artificial orthography. *J. Exp. Psychol. Gen.* 146 (6), 826–858.
- Taylor, J.S.H., Rastle, K., Davis, M.H., 2014. Distinct neural specializations for learning to read words and name objects. *J. Cognit. Neurosci.* 26 (9), 2128–2154.
- Taylor, M.J., Donner, E.J., Pang, E.W., 2012. fMRI and MEG in the study of typical and atypical cognitive development. *Neurophysiol. Clin. Clin. Neurophysiol.* 42 (1–2), 19–25.
- Torgesen, J.K., Rashotte, C.A., Wagner, R.K., 1999. TOWRE: Test of Word Reading Efficiency (Pro-Ed Austin, TX).
- Van Atteveldt, N., Formisano, E., Goebel, R., Blomert, L., 2004. Integration of letters and speech sounds in the human brain. *Neuron* 43 (2), 271–282.
- van Atteveldt, N.M., Formisano, E., Goebel, R., Blomert, L., 2007. Top-down task effects overrule automatic multisensory responses to letter-sound pairs in auditory association cortex. *Neuroimage* 36 (4), 1345–1360.
- van Atteveldt, N., Roebroek, A., Goebel, R., 2009. Interaction of speech and script in human auditory cortex: insights from neuro-imaging and effective connectivity. *Hear. Res.* 258 (1–2), 152–164.
- Varga, V., Tóth, D., Csépe, V., 2020. Orthographic-phonological mapping and the emergence of visual expertise for print: a developmental event-related potential study. *Child Dev.* 91 (1), e1–e13.
- Vigliocco, G., 2000. Language processing: the anatomy of meaning and syntax. *Curr. Biol.* 10 (2), R78–R80.
- Vinckier, F., Dehaene, S., Jobert, A., Dubus, J.P., Sigman, M., Cohen, L., 2007. Hierarchical coding of letter strings in the ventral stream: dissecting the inner organization of the visual word-form system. *Neuron* 55 (1), 143–156.
- Wechsler, D., 2008. Wechsler Adult Intelligence Scale–Fourth Edition (WAIS-IV), vol. 22. NCS Pearson, San Antonio, TX, p. 498.
- Wilson, S.M., Bautista, A., McCarron, A., 2018. Convergence of spoken and written language processing in the superior temporal sulcus. *Neuroimage* 171, 62–74.
- Xu, W., Kolozsvári, O.B., Monto, S.P., Hämäläinen, J.A., 2018. Brain responses to letters and speech sounds and their correlations with cognitive skills related to reading in children. *Front. Hum. Neurosci.* 12, 304.
- Xu, W., Kolozsvári, O.B., Oostenfeld, R., Leppänen, P.H.T., Hämäläinen, J.A., 2019. Audiovisual processing of Chinese characters elicits suppression and congruency effects in MEG. *Front. Hum. Neurosci.* 13, 18.
- Yonelinas, A.P., 2013. The hippocampus supports high-resolution binding in the service of perception, working memory and long-term memory. *Behav. Brain Res.* 254, 34–44.
- Žarić, G., Fraga González, G., Tijms, J., van der Molen, M.W., Blomert, L., Bonte, M., 2014. Reduced neural integration of letters and speech sounds in dyslexic children scales with individual differences in reading fluency. *PLoS One* 9 (10), e110337.