universität
wien

MASTERARBEIT / MASTER'S THESIS

Titel der Masterarbeit / Title of the Master's Thesis

„Look what I'm saying! Neural correlates of lipread vs. text-based phonetic recalibration in dyslexic adults"

verfasst von / submitted by

Anna Palmann, BSc BA

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of

Master of Science (MSc)

Wien, 2021 / Vienna 2021

Studienkennzahl lt. Studienblatt / degree programme code as it appears on the student record sheet:	UA 066 840
Studienrichtung lt. Studienblatt / degree programme as it appears on the student record sheet:	Masterstudium Psychologie UG2002
Betreut von / Supervisor:	Univ.-Prof. Dr. Frank Scharnowski, MSc
Mitbetreut von / Co-Supervisor:	

Acknowledgements

I would first and foremost like to thank my supervisor Linda Romanovska for her invaluable advice, the skills she taught me, and all the mails that went back and forth. Thank you for making my research internship and the writing of my thesis a pleasant experience.

I would further like to thank Milene Bonte for letting me join her lab for a little while and her lab meetings for a little longer. Many thanks also to the people in the Brain and Language group: you are an inspiring bunch - both, academically and also on a personal level. Thank you for reminding me that research is based on collegiality and curiosity. Many thanks also to Frank Scharnowski, my supervisor in Vienna, for letting me do my thing abroad while at the same time always being there when I needed support or advise.

It's both, a bit of a blessing and a curse to have been a linguist among psychologists and a psychologist among linguists ever since the beginning of my studies. I would like to thank all the professors who taught me to adopt an interdisciplinary mindset and the importance of looking at a phenomenon from a healthy distance, and always from different angles.

Many thanks also to my friends - far away and very close. My Kleeblatt family for more than a decade of friendship and support. Thanks to Pep for keeping me company in the computer room and proofreading this thesis, and to the people the Landhuis or at any other place where I felt like it was the right time to work on my thesis - be it in the Netherlands, Austria, or Italy.

תודה רבה

to Nitzan for always listening to my thoughts and ideas - in the early morning, the late evening, and at any point in between. For not leaving the room even though I was hammering in the keyboard, and for pouring the exact right amount of plant-based milk in my coffee and handing it to me with no comment but a supportive smile (knowing that I would always accept a cup).

Finally, I would like to thank my parents for teaching me genuine curiosity towards all things and planting the seed of my fascination for language. For being interested in what I am working on at the moment and actually trying to understand it and its value. And for supporting me always and everywhere without thinking twice about it.

Abstract

Everyday life is full of ambiguous speech sounds, and contextual information has been found to facilitate their perception and disambiguation. In individuals with developmental dyslexia, phonological processing and letter-speech sound integration are impaired, leading to difficulties in reading. In this fMRI study, dyslexic adults were exposed to a phonetic recalibration paradigm with ambiguous speech sounds. Additionally, participants were presented simultaneously with either written text or video recordings of lip movements, which served as disambiguating information. The strength of perceptual shift (recalibration) was calculated comparing the two conditions hypothesizing that lip movements would serve as better disambiguating stimulus than text in this specific sample. There was however a significant recalibration effect in both conditions and no significant difference in strength of effect between the two. As hypothesized, brain areas associated with reading and audio-visual integration, such as the visual, auditory, and motor cortex were active in both conditions. Further, there were stimulus-dependent differences in neural activation, with the text-based condition eliciting stronger activation in the posterior occipito-temporal cortex and visual word form area (VWFA) while the lipread condition lead to a broader activation cluster in the occipital cortex including fusiform face area (FFA). The text-based condition further elicited activation in the left inferior frontal gyrus (IFG). Given the significant recalibration effect for both, written text and lip movements and the involvement of the left IFG during text processing, it can be concluded that participants successfully employed compensation strategies during the integration of written text and speech.

Keywords: dyslexia, speech perception, audiovisual integration, phonetic recalibration, fMRI

Table of Contents

"How do I know what I think until I see what I say?"

- E.M. Forster

Introduction

Developmental dyslexia (DD) is a learning disorder characterized by deficits in reading and spelling (Rüsseler et al., 2018). The *Diagnostic and Statistical Manual of Mental Disorders* (DSM-5) and the *International Statistical Classification of Diseases and Related Health Problems* (ICD-10), both major diagnostic tools, note as main definition criterion a lower performance than average on standardized reading tests, which subsequently impacts school achievement and the use and amount of reading in everyday life (Habib, 2020). In the DSM-5, DD is classified as heritable neurodevelopmental condition with an early onset persisting into adulthood (Snowling et al., 2020). Its prevalence is irrespective of IQ and the deficits cannot be explained through general learning difficulties (Snowling et al., 2020), other pathologies such as sensory or neurological damage, or poor educational opportunities (Rüsseler et al., 2018). Di Folco et al. (2020) compared the different definitions for developmental dyslexia in DSM-5 and ICD-10 and found that prevalence rate is highly impacted by the exact definition applied together with specific criteria, such as mother tongue or socioeconomic background. Therefore, the prevalence of DD in children with respect to the overall population of the same age may vary from 5% to 15% (Habib, 2020).

Developmental dyslexia comprises impairments in reading fluency, accuracy, and comprehension (Nergård-Nilssen & Hulme, 2014; Romanovska et al., 2019). Phonological awareness, the ability to perceive and manipulate sounds in one's language, underlies these functions and is crucial for overall successful reading (Ozernov-Palchik & Gaab, 2016). The fact that dyslexic readers perform poorer in general speech sound categorization tasks (Habib, 2020) hints towards deficits in phonological processing (Rüsseler et al., 2018). Another process crucial for fluent reading is the integration of letters and speech sound, also referred to as grapheme-phoneme conversion. Dyslexic readers show deficits in the cognitive abilities associated with general audio-visual processing (Rüsseler et al., 2018), which is necessary for mastering the association between spoken and written language (Habib, 2020). More precisely, it might be the case that dyslexic readers are able to associate but not successfully integrate pairs of letters and sounds (Blomert, 2011).

Dyslexia and the brain

Dyslexia is associated with atypicalities in the reading network of the brain (Ozernov-Palchik & Gaab, 2016). This neural reading circuit comprises:

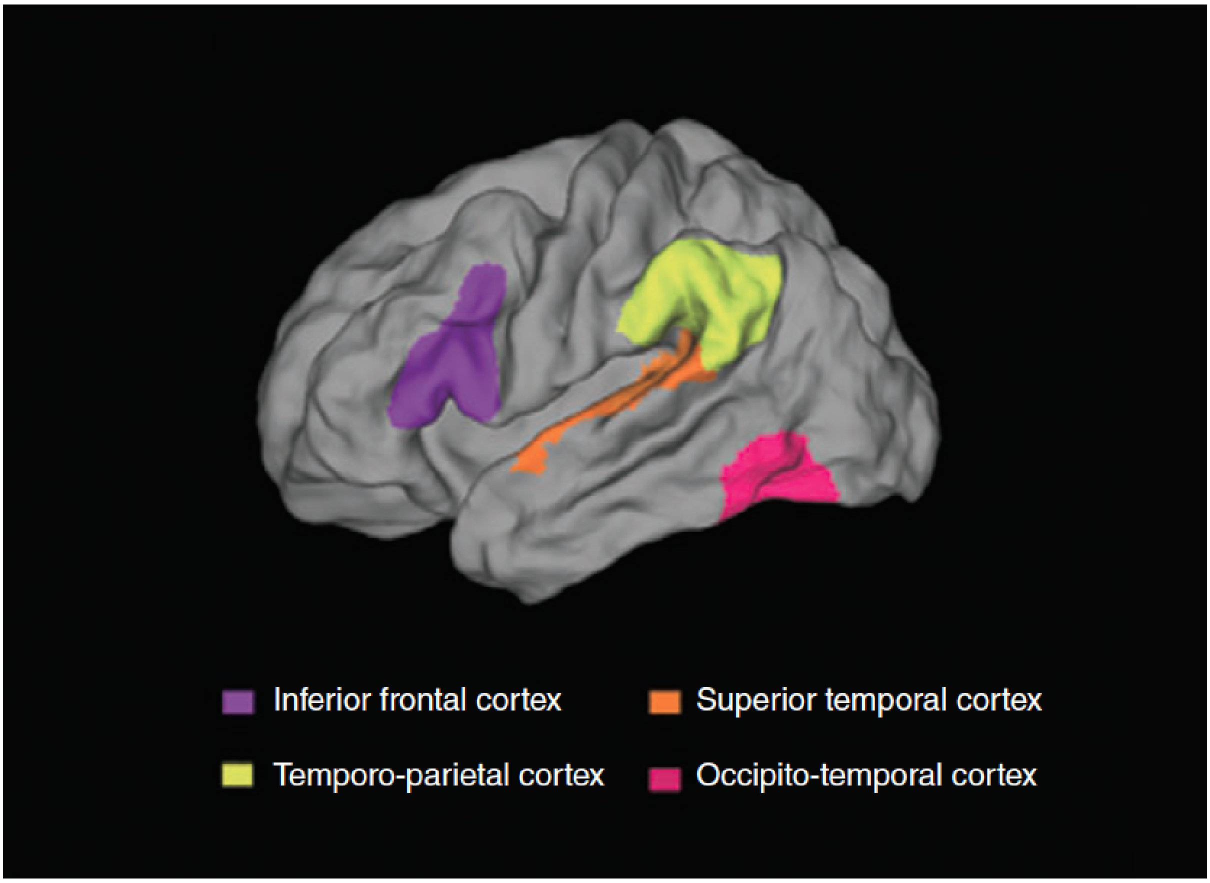
- the dorsal temporo-parietal network (superior temporal, supramarginal, and angular gyrus) associated with speech production (Mirman et al., 2015) including the integration of phonology and orthography and word analysis (Ozernov-Palchik & Gaab, 2016);

- the ventral occipito-temporal circuit (lateral extrastriate, fusiform, and inferior-temporal regions including the visual word-form area) specialized for speech recognition and comprehension (Mirman et al., 2015), especially also print and rapid word processing (Ozernov-Palchik & Gaab, 2016);
- and the anterior inferio-frontal circuit (including Broca's area) associated with a number of different less clearly defined functions in reading, such as phonological processing, speech planning, lexical access and semantics, as well as speech comprehension and more general cognitive functions like inhibition and attention (Ozernov-Palchik & Gaab, 2016).

Dyslexic readers display hypoactivation in all three of these networks, especially in the left hemisphere as depicted in figure 1, accompanied by reduced functional connectivity between key areas of these networks including superior temporal gyrus, fusiform gyrus, and occipital gyrus (Ozernov-Palchik & Gaab, 2016).

Figure 1

Reading network in the brain impaired in dyslexia



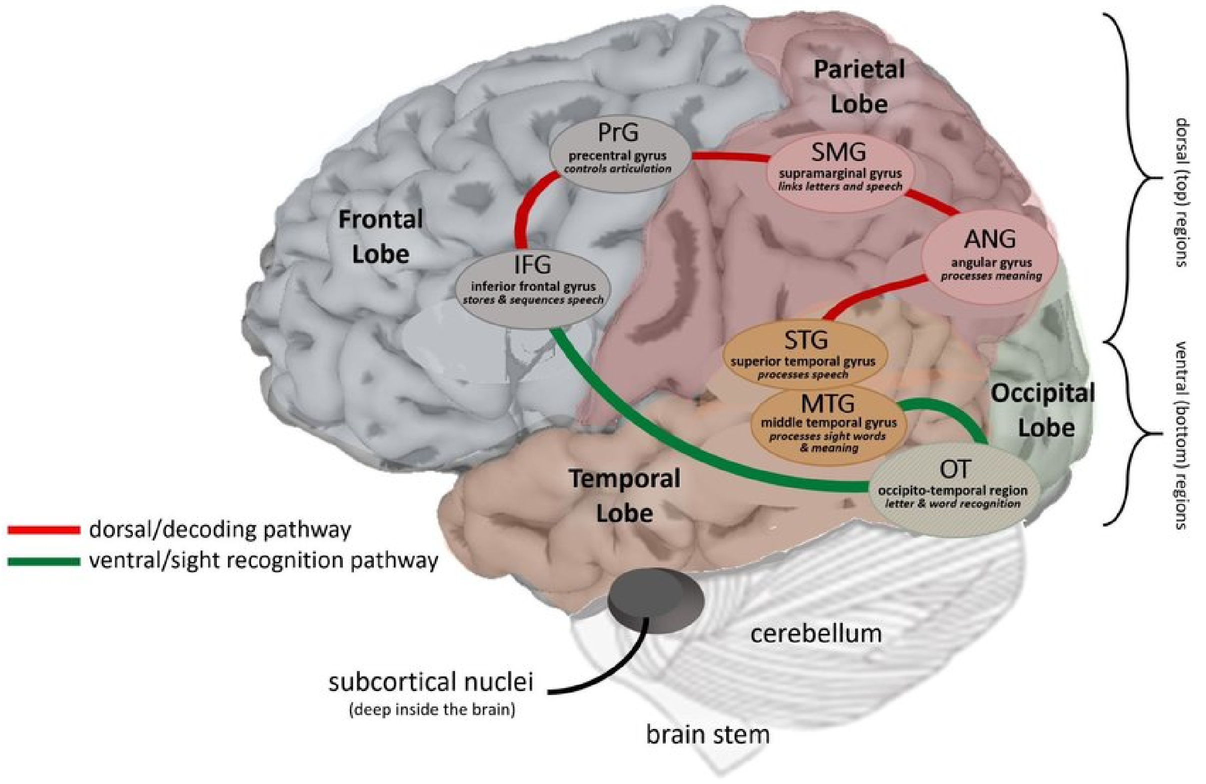
Note. This figure (taken from Ozernov-Palchik & Gaab, 2016) displays brain regions important for reading associated with atypical function or structure in people with dyslexia.

Phonological processing in dyslexic readers is associated with lower activity in the left occipito-temporal and temporo-parietal regions, and the bilateral cerebellum (Raschle et al., 2011). This was measured bx Raschle et al. (2011) using a first-sound matching task, where participants were required to indicate whether two words start with the same or a different sound. On the other hand, orthographic processing deficits in individuals with dyslexia are more associated with hypoactivation in several posterior dorsal and anterior frontal regions (Ozernov-Palchik & Gaab, 2016). Brain structural differences are noticeable early in development when comparing children at risk of developing dyslexia due to a family history of DD, and those without. In the aforementioned study, Raschle et al. (2011) were able to show reduced gray matter in children at risk of dyslexia in the left occipito-temporal regions, bilateral temporo-parietal regions, left fusiform gyrus, and right lingual gyrus.

Further findings hint towards different strategies in reading when comparing typically and non-typically evolving readers. These differences are also associated with diverging neural activation patterns. In general, it is possible to differentiate between two neural pathways involved in reading: the dorsal and the ventral pathway (see figure 2).

Figure 2

Dorsal and ventral reading pathways in the brain



Note. This figure (taken from Kearns et al., 2019) displays the dorsal decoding (red) and the ventral recognition (green) pathways in the brain that are active during reading.

The ventral automated text processing pathway including the left fusiform area, is employed by unimpaired readers in most cases (Romanovska et al., 2021). The dorsal reading pathway on the other hand, is used for systematic assessment of phoneme-grapheme connections and more explicit mappings of letters and speech sounds, for example when words are not familiar (Kearns et al., 2019). Dyslexic individuals have been found to mainly use the dorsal pathway during reading (Shaywitz & Shaywitz, 2008).

When directly comparing brain regions involved in reading development of dyslexic and non-dyslexic children, Yamada et al. (2011) found that typical reading development is associated with an initial bilateral recruitment of the dorsal reading network and a subsequent disengagement of the right hemisphere, while non-typical readers show reduced recruitment of bilateral regions, followed by a compensatory involvement of frontal regions - especially pronounced after a reading intervention. In terms of compensatory mechanisms, dyslexic readers may show reduced activation in the posterior reading system during phonological analysis, but at the same time there is increased activation in the left and right inferior frontal gyrus, and the right occipito-temporal areas (Shaywitz & Shaywitz, 2008). This points to the development of a more posterior and medial, rather than an anterior and lateral reading system in dyslexic individuals (Shaywitz & Shaywitz, 2008).

Audio-visual integration of speech

Reading development is essentially multi-sensory in nature (Blomert, 2011), and as previously established the process of visual word encoding (i.e. reading) requires the decoding of visual and auditory information in interaction. Everyday life is full of ambiguous speech sounds in form of differences between dialects, vague pronunciations, or background noise (Keetels et al., 2016). Contextual information is known to facilitate speech perception and disambiguation (Baart et al., 2012). As an example, gestures facilitate understanding of spoken words, since speech comprehension frequently relies on a combination of auditory and visual information (Rüsseler et al., 2018). But also lexical knowledge, lipreading (Bertelson et al., 2003), or written text (Keetels et al., 2018) have been found to have an impact on the disambiguation of spoken word. While the connection between speech and lip movements is rooted in biology and can be explained in evolutionary terms (Keetels et al., 2016), the association of speech and written text is an arbitrary one and rooted in culture, thus only acquired later in life with the onset of formal reading instruction (Liberman, 1992).

Phonetic recalibration

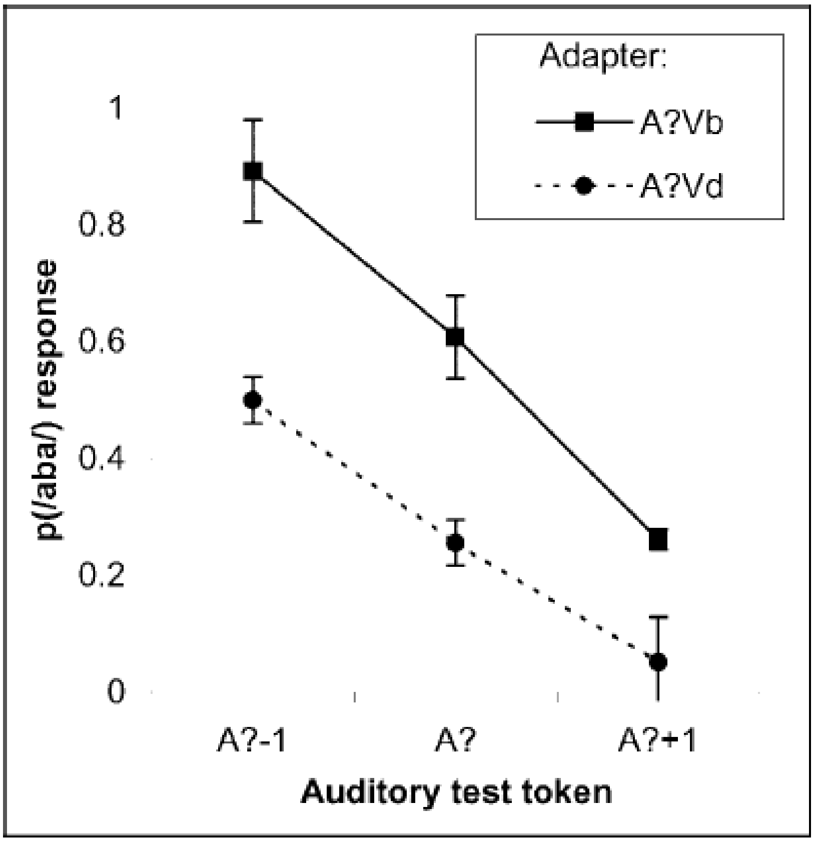
The influence of cross-modal contextual factors, such as lip movements or gestures on the perception of speech sounds, has been prominently depicted with the *McGurk Effect*

(McGurk & MacDonald, 1976). In an audiovisual exposure paradigm, participants were presented with audio recordings of repeated utterances of the syllable "ba" combined with video recordings of lip movements of the syllable "ga", which led to the perception of the sound "da".

Bertelson et al. (2003) could show as well that lipread information can influence and recalibrate auditory speech identification. They reported a so-called *McGurk aftereffect*, where - even after some time delay - subjects selectively adapted to perceive a certain sound, following exposure to audiovisual stimuli consisting of an ambiguous sound together with disambiguating visual information. This phenomenon has been entitled *phonetic recalibration effect* (Baart et al., 2012) since the perception of phoneme boundaries is adjusted according to an additional visual stimulus. In their study, Bertelson et al. (2003) used a lipread recalibration paradigm manipulating the ambiguity of the auditory speech information. They presented participants with various combinations of audio and video recordings of a person pronouncing the sounds "aba" and "ada". First, the auditory signals were changed resulting in a 9-token continuum. This continuum consisted of the clear "aba" sound on one end and the clear "ada" sound on the other, and between them seven ambiguous sounds that were created by morphing the two. From this continuum, for each participant the most-ambiguous defined speech sound "a?a" was determined individually in a pre-test. During subsequent exposure trials, this ambiguous sound was combined with either "ada" or "aba" video recordings. After eight exposure trials of the same condition ("aba" or "ada"), participants again underwent an auditory post-test, where they were presented with their most ambiguous sound, as well as with the neighbouring sounds on both sides of the continuum. In a total of six trials, they were then asked to indicate whether they perceived the sound as "aba" or "ada". Participants did indeed show higher "aba" responses in the post-test after exposure to "aba" lip movements in the exposure trials. Accordingly, they showed higher "ada" responses in the post-test after exposure to "ada" lip movements in the exposure trials (see figure 3). Thus, a recalibration effect was present.

Figure 3

Results from the post-test in the study by Bertelson et al. (2003)



Note. The figure (taken from Bertelson et al., 2003) displays the proportion of /aba/ judgements (y-axis) to the three post-test sounds (x-axis) after exposure to audiovisual trials with /aba/ (solid line) or /ada/ (dotted line) as disambiguating text.

In order to make sure that this effect was not due to selective speech adaptation, i.e. the phenomenon that after repeated exposure a certain speech token is less likely to be reported in a subsequent identification task (Bertelson et al., 2003), the authors conducted another experiment, in which they also presented congruent non-ambiguous auditory stimuli during exposure blocks. The inclusion of these adaptation blocks further served the purpose of controlling for a possible response bias induced by mere exposure to the stimuli. Thereby, it was possible to rule out the possibility that participants were for example choosing "aba" as response simply because they saw the corresponding lip movement or text. In the congruent non-ambiguous (adaptation) condition, there was in fact a selective speech adaptation effect noticeable, so that participants showed less "aba" responses during "aba" exposure (and the same for "ada"). As previously established, this was the opposite case for the ambiguous (recalibration) condition. The latter is the one used in the first experiment, therefore selective speech adaptation can be ruled out as an explanation.

Keetels et al. (2016) adapted the recalibration paradigm from Bertelson et al. (2003) in order to investigate text-based phonetic recalibration effects. They were able to show that also disambiguating orthographic information (text) directs participants to adjust their phoneme boundary accordingly. In their study, audiovisual exposure blocks consisted of an ambiguous speech sound combined with the text "aba" or "ada" for disambiguation (recalibration condition), or a non-ambiguous speech sound combined with the congruent text (adaptation condition). This was followed by auditory-only test trials, in which participants were asked in a forced-choice paradigm to indicate whether the speech sound they heard was "ada" or "aba". For the ambiguous condition, there was a significant recalibration effect (responses significantly in line with the disambiguating text) while for the non-ambiguous condition there was the opposite effect present (less responses in line with the text), which indicates a selective speech adaptation effect (Keetels et al., 2016). The fact that there were both, significant recalibration and adaptation effects, shows that the text-based recalibration paradigm used by Keetels et al. (2016) induces indeed reliable perceptual shifts.

Van Linden and Vroomen (2007) directly compared lipread and text-based phonetic recalibration and found that the effects were similar in size, build-up, and dissipation rate. However, comparing lexical (in form of spoken word and not text) and audiovisual (lipread) recalibration, Ullas, Formisano, et al. (2020) found effects to be lower for lexical than for audiovisual cues or a combination of the two. This points towards a higher recalibration effect when lip movements rather than text serve as disambiguating stimuli. Comparing both, lipread and text-based recalibration effects in adults with and without developmental dyslexia, Keetels et al. (2018) found as well that overall, lipread recalibration effects were stronger than text-based ones. The effect was stronger for unimpaired readers than for DD subjects, so that the latter only showed lipread but no text-based aftereffects. Furthermore, reading fluency negatively correlated with the difference between lipread and text-based recalibration effects (Keetels et al., 2018). These findings go in line with Baart et al. (2012), who found the lipread recalibration effect to be in the normal range for dyslexic readers. Interestingly however, Romanovska et al. (2019) found similar text-based recalibration effects for dyslexic and typically reading children. In the dyslexic but not the other group, the strength of the recalibration effect correlated with the strength of perceptual adaptation effects measured in a separate adaptation task. There was further a significant link between recalibration strength and phoneme categorization skills in both groups (Romanovska et al., 2019).

Neural correlates of phonetic recalibration

The processing of audiovisual speech is associated with activation in a broad network in the brain, with a noticeable left-over-right asymmetry. Areas involved are the middle and posterior sulcus, Broca's area, dorsolateral prefrontal cortex (DLPFC), superior precentral sulcus, supramarginal gyrus, angular gyrus, intraparietal sulcus, inferior frontal gyrus (IFG), and insula (Kilian-Hütten et al., 2017). Because of this broad activation pattern, multisensory processing mechanisms are assumed to be in play rather than initially unimodal and then higher-order integration of auditory and visual information (Kilian-Hütten et al., 2017). There are two approaches to explain how this integration comes about. Ones is based on the correlation of temporal dynamics between auditory and visual signal while the other assumes that the visual signal helps to disambiguate the auditory information in a more complementary manner (Kilian-Hütten et al., 2017).

Using functional magnetic resonance imaging (fMRI) and multi-voxel pattern analysis (MVPA), Kilian-Hütten et al. (2011) investigated which brain areas are active during the perception of ambiguous speech sounds in a lipread recalibration paradigm. They found broad activation in early auditory cortical areas, such as in the superior temporal cortex including Heschl's gyrus (HG), planum temporale (PT), superior temporal gyrus (STG) and superior temporal sulcus (STS). Other activated areas were the middle temporal gyrus, insula, inferior parietal cotex (IPC), inferior frontal cortex (IFC), and the supramarginal gyrus. MVPA further showed a distinct activation pattern for "aba" and "ada" exposure conditions in the posterior bank of Heschl's gyrus and sulcus, as well as in the anterior planum temporale.

Bonte et al. (2017) looked at brain activation during exposure blocks in a text-based recalibration task. Using multivariate fMRI, the main changes in activation were found in the posterior superior temporal cortex and early auditory regions, which goes in line with the results from Kilian-Hütten et al. (2011). This shows that not only video recordings of lip movements, but also disambiguating textual information induce a significant and clearly identifiable shift in the perception of an ambiguous sound, also visible in changes in the neural activation pattern. The authors were further able to retrieve the text-induced perceptual interpretations from the activity patterns in the posterior superior temporal cortex. By applying psychophysiological interaction analyses, they found functional connectivity of this area with the inferior parietal lobe (IPL) during text-speech pairing. This indicates that the IPL is involved in establishing the shift in perception during phonetic recalibration.

Ullas, Hausfeld, et al. (2020) presented disambiguating lexical (audio recordings) vs. audiovisual (lip movements) stimuli to non-impaired readers and investigated the resulting changes in neural activation. Both forms of cues were found to be associated with a network comprising the temporal and parietal cortex, the insula, and the motor areas. More precisely, the active areas were the bilateral HG and PT (early auditory areas), the STG and STS (associated with phoneme and syllable processing), the left IPL (involved in processing of audiovisual information), and the right insula. As only major difference between conditions, the audiovisual task unsurprisingly elicited additional activation in the occipital (visual) cortex (Ullas, Hausfeld, et al., 2020) .

Rüsseler et al. (2018) investigated audiovisual integration in dyslexic and non-dyslexic adults using event-related fMRI and an independent component analysis approach. The authors applied a passive exposure paradigm, during which congruent and incongruent audio-visual stimuli were presented to the participants. Stimuli consisted of audio recordings of a female native German speaker pronouncing disyllabic nouns, as well video recordings of the same speaker while uttering the word. In the congruent condition the video recording showed the speaker uttering the same sound sequence that was presented acoustically, while in the incongruent condition the presented video was taken from another word, so that the visual and the acoustic component of the stimulus did not match. There was less neural activation for incongruent stimuli in dyslexic adults as compared to typical readers. Accordingly, there was higher activation for congruent stimuli in non-dyslexic participants. Overall, less neural activation in typical audio-visual processing and integration areas (e.g. STS) and face processing areas (fusiform gyrus, occipital gyrus) was found for the dyslexic group indicating a general impairment of audiovisual processing in readers with DD (Rüsseler et al., 2018).

In a most recent study, Romanovska et al. (2021) used the same text-based recalibration paradigm as Keetels et al. (2018) comparing dyslexic and non-dyslexic children. While there was a significant effect for both groups in the recalibration task outside of the scanner, the recalibration effect during the fMRI scan was reduced for both groups and was not significant. The authors looked at brain activation during the exposure phase and found a similar activation pattern for both groups in a broad bilateral network in the superior temporal cortex including occipital cortex, left fusiform areas, bilateral superior temporal gyrus, and frontal and parietal areas. Further, they also found an interaction effect between behavioural and neural data, where activation in the bilateral STG correlated negatively with participant's performance in a task measuring grapheme-phoneme conversion. This means that the lower the scores for grapheme-phoneme conversion, the higher the activation in this area. These results are most likely due to less automated processing, since the STG is involved in explicit letter-speech sound mapping rather than automated reading. A region of interest (ROI) analysis further revealed reduced activation in the left fusiform area despite similar recalibration effects, for dyslexic as compared to non-dyslexic children. Activation in this area was linked to better reading and phonological skills, which explains its lower activation in children with dyslexia. These differences in brain activation despite comparable behavioural recalibration effects may also be explained by the previously mentioned different recruitment of the dorsal, more explicit and ventral, more automated reading systems in readers with and without dyslexia.

The present research

The goal of this study is to investigate the neural mechanisms underlying audio-visual integration of ambiguous speech with text and with lip movements in dyslexic adults, using a recalibration paradigm and fMRI. So far, no study has directly compared lipread and text-based recalibration in dyslexic adults, taking into account associated neural activation patterns. It seems therefore worthwhile to compare these two types of recalibration and see whether there is a difference both, behaviourally in form of a recalibration effect, as well as in neural activation during the exposure phase. A possible association between recalibration effects and neural activity in form of an interaction effect will also be investigated.

Furthermore, it is one of the aims of this study to acknowledge and highlight the multi-sensory nature of speech processing by taking into account factors that also play a role in more naturalistic settings. The focus on different types of audiovisual integration and their neural correlates, will enable to better understand and disentangle their underlying mechanisms and possible associations with two developmentally separate pathways in individuals with dyslexia. These findings could subsequently be applied in order to detect children at risk of dyslexia and/or provide them with well-tailored support during reading instruction. By applying a well-established paradigm similar to the ones used by Keetels et al. (2018) and Romanovska et al. (2021), we assure that the comparison of the two conditions is valid. Similar to the previously described studies, the focus will lie on the neural activation during exposure blocks, not during post-test trials. The responses provided during the post-test trials will serve the purpose of assessing the behavioural recalibration effect.

Based on the findings by Keetels et al. (2016) and Ullas, Hausfeld, et al. (2020), we hypothesize there to be a stronger behavioural recalibration effect for the lipread condition compared to the text-based one. This is based on the assumption that impaired phoneme-grapheme conversion in dyslexic readers will lead to worse performance when text serves as disambiguating information. In terms of neural activation, we assume there to be a similar activation pattern for both conditions (cf. Romanovska et al., 2021). Based on findings from studies using a similar recalibration paradigm (Bonte et al., 2017; Kilian-Hütten et al., 2011; Romanovska et al., 2021; Ullas, Hausfeld, et al., 2020), we hypothesize to find activation in

- the superior temporal cortex including HG, PT, and temporoparietal junction (TPJ)
- the occipital cortex
- the parietal cortex, as well as insula and motor areas

We further expect behavioural recalibration effects to be correlated with neural activation in selected brain areas. In line with (Ullas, Hausfeld, et al., 2020) who found a correlation for the lipread but not the lexical condition, we assume this correlation effect to also be present for the lipread, but not in the text-based condition in our study. The direction of the correlation as well as the question which specific brain areas, will not be defined beforehand so that the correlation analysis will be exploratory in nature.

Methods

Participants

Participants were ten adult native Dutch speakers (7 females, mean age: 26.3 ± 4.9) recruited via social media and fliers. All had normal or corrected to normal vision, no reported hearing problems, and no history of other diagnosed developmental or neurological disorders. All but one participant were right-handed. All participants except for one had been given an official dyslexia diagnosis in childhood, on average at the age of 11 years. Informed consent was given in accordance with the declaration of Helsinki and ethical approval was obtained via Maastricht University's Ethics Review Committee Psychology and Neuroscience (ERCPN). Participants received gift vouchers as reward for their participation.

Assessment of reading and phonological skills. Reading and phonological skills were assessed using the *Eén-Minuut-Test* (Brus & Voeten, 1991), a screening instrument for dyslexia specifically tailored for Dutch speaking population. In this test, single-word reading for real words is assessed by asking participants to read out loud as many words as possible from a given list within the time span of one minute. Reading fluency scores are calculated by subtracting the number of mistakes from the total number of read words. Further, a Spoonerism task similar to the one described by Nergård-Nilssen and Hulme (2014) was applied. This task measures the ability to segment and manipulate individual speech sounds by presenting participants with two real words at a time and instructing them to repeat the words after having swapped the initial sounds. In total, the task consisted of 20 real-word pairs. Scores were calculated using overall time to complete all 20 items, and accuracy, defined as the number of correct swaps.

Table 1

Age and results from the behavioural reading tests

	Mean	SD	Range
Age	26.3	5.25	18-35
Eén-Minuut-Test (fluency)	75.8	11.3	58-93
Eén-Minuut-Test (accuracy)	74.9	11.1	57-91
Spoonerism task	121.83	44.94	65-212.7

Note. This table shows the mean, standard deviation (SD), and range for participants' age and scores in the Eén-Minuut-Test and the Spoonerism task

Stimuli

Both, speech and visual stimuli for the recalibration task were the same as used by Bertelson et al. (2003).

Speech stimuli consisted of recordings of a male native Dutch speaker pronouncing the non-words "aba" and "ada", lasting 650 ms each. These sounds were used to create a 9-step continuum ranging from clear /aba/ to clear /ada/ sounds, by changing the second formant (F2) in 8 steps of 39 Mel using the *Praat speech editor* (Boersma, 2001). The resulting ambiguous sounds between clear /aba/ and clear /ada/ were used in the pre-test. The auditory stimuli were presented binaurally through MR compatible headphones (Sensimetrics, model S14, https:www.sens.com) at a comfortable listening volume.

Visual stimuli for the text-based recalibration consisted of the non-words "aba" and "ada" presented in white at the center of a black screen in Times New Roman font (font size 50). Text stimulus presentation duration was 1 s. Visual stimuli for the lipread recalibration consisted of video recordings of a male person pronouncing 'aba' or 'ada'. The videos showed the face of the speaker from the forehead to the chin and had a duration of 720 ms. Auditory and visual stimuli were presented using *Presentation software* (Version 18.0, Neurobehavioral Systems, Inc., Berkeley, CA, www.neurobs.com). All auditory and visual stimuli were presented in a 900 ms silent gap between fMRI volume acquisitions.

Experimental design

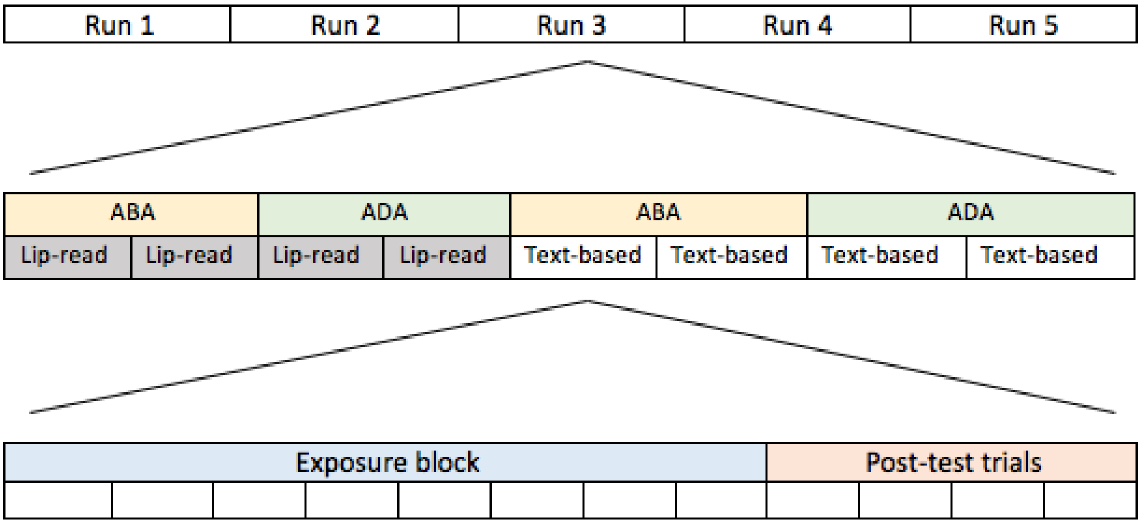
Each participant visited the lab for two separate sessions, one in which the recalibration task was performed on the PC, and one scanning session. In both sessions, participants underwent a pre-test, which served the purpose of identifying their individual most ambiguous speech sound. This study will focus solely on the results from the recalibration task performed during the scanning session.

Recalibration task

For the recalibration task, we adopted a study design similar to the one used by Keetels et al. (2016). Figure 4 shows an overview of the experimental design. The recalibration task in the scanner consisted of 5 runs with 8 blocks each (4 lipread, 4 text-based), resulting in a total of 40 blocks. In each run, the lipread blocks and the text-based blocks were presented in a grouped manner, one after the other. In each run, the participants were exposed to 2 “aba” and 2 “ada” text blocks, as well as 2 ‘aba’ and 2 ‘ada’ video blocks. For the 5 runs, this resulted in a total of 20 blocks per condition (text-based vs. lipread) with 50% of the blocks always being "aba" and the other half "ada". These "ada" and "aba" blocks were presented in pseudo-randomized order making sure that neither was presented more than twice in a row (figure 4 therefore depicts only an example of a possible combination of the "aba" and "ada" blocks). A single text-based or lipread block consisted of 8 audiovisual exposure blocks (auditory stimulus combined with either video or text) followed by 4 auditory-only post-test trials. The inter-trial interval between the audiovisual stimuli was 2 s, resulting in a total duration of 16 s per exposure block. Exposure blocks were separated from post-test blocks by a jittered rest period of 8 - 12 seconds, with an average length of 10 s. During this jittered rest period, which served as baseline for subsequent statistical analyses, a white fixation cross was presented on a black screen.

Figure 4

Experimental design for the scanning session

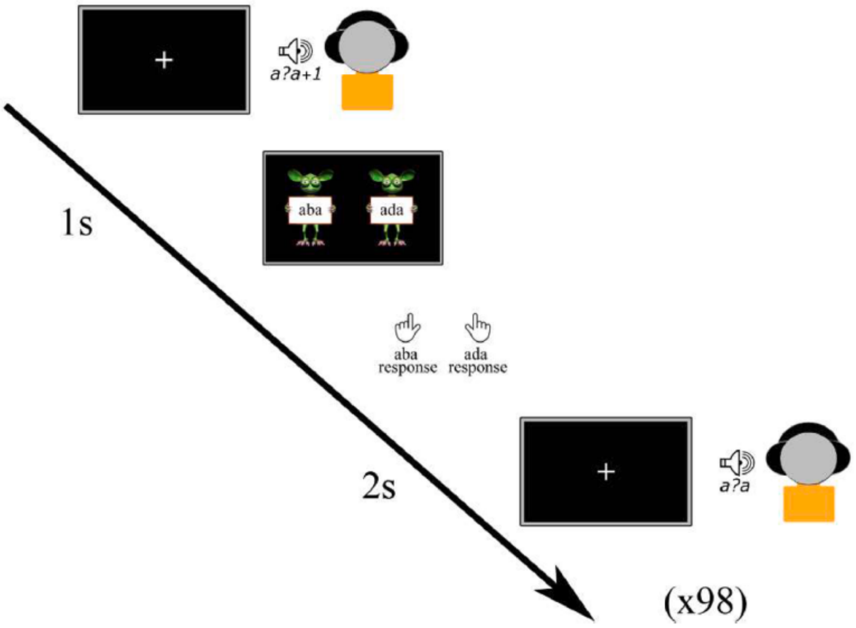


Pre-test. In a pre-test, we identified the most ambiguous speech sound for each participant individually. This sound, denoted as /a?a/, was used subsequently in the audio-visual exposure blocks and post-test trials of the recalibration task. In the post-test trials, additionally the flanking sounds on both sides of the continuum /a?a/-1 (closer to /aba/) and /a?a/+1 (closer to /ada/) were used.

During the pre-test, participants were presented with all 9 sound tokens on the /aba/-/ada/ continuum 98 times in a randomized order, with the ambiguous sounds presented more often than the clear /aba/ and /ada/ ones. Participants were instructed to listen to each sound and then indicate whether they heard /aba/ or /ada/ by pressing the left or right innermost button on a button box with their left/right index fingers, respectively. During the presentation of the speech sound, participants viewed a black screen with a white fixation cross, followed 1 s later by a response cue consisting of illustrations of monsters holding the text stimuli "aba" on the left, and "ada" on the right side (see figure 5). Two seconds after the response was given by the participant, the next stimulus was presented. The most ambiguous speech sound was then determined based on the proportion of "aba" responses to each token, and was defined as the sound with an /aba/ or /ada/ response closest to 0.5, indicating the phoneme boundary (Romanovska et al., 2019). In total, the pre-test lasted approximately 5 minutes.

Figure 5

Pre-test

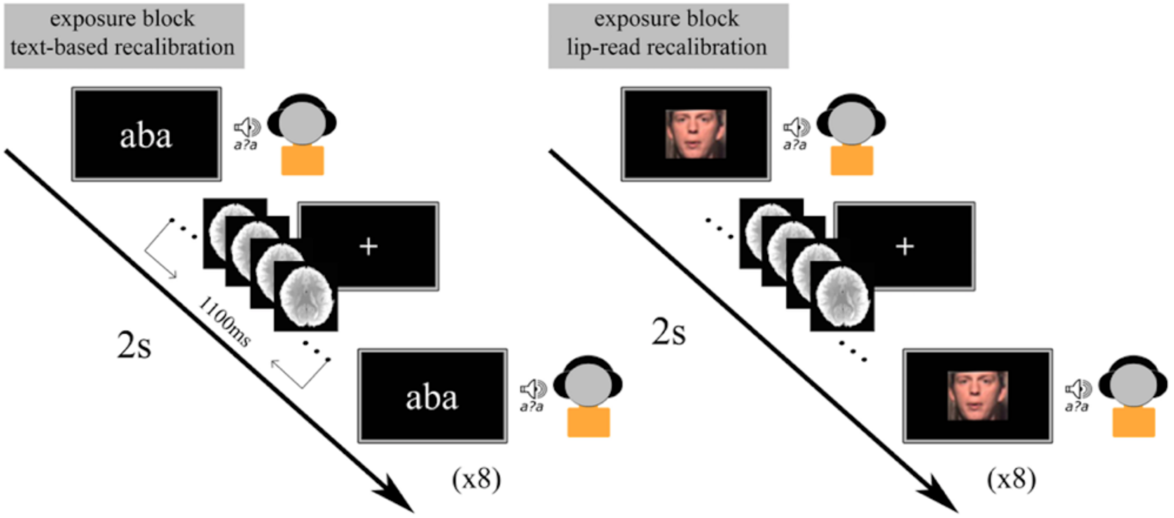


Note. This figure (taken from Romanovska et al., 2021) displays the experimental setup for the pre-test, where the participant is presented with the 9 sound stimuli to determine the individual most ambiguous sound by indicating with a button-press whether the heard sound was "aba" or "ada".

Exposure blocks. During audio-visual exposure blocks, participants were presented with 8 trials of their individually defined most ambiguous speech sound /a?a/ in combination with either the disambiguating text or the lip movement stimulus for either "aba" or "aba" (see figure 6). During the exposure phase, participants were instructed to pay close attention to the visual and auditory stimuli without providing a response. Auditory and visual stimuli were presented simultaneously, the auditory stimulus had a duration of 650 ms, while the text stimulus lasted 1s and the video stimulus 720 ms. Therefore, overall trial duration was 1 s in both conditions, where auditory and visual stimuli were presented simultaneously with a stimulus-onset time of 0 and had different duration.

Figure 6

Exposure blocks

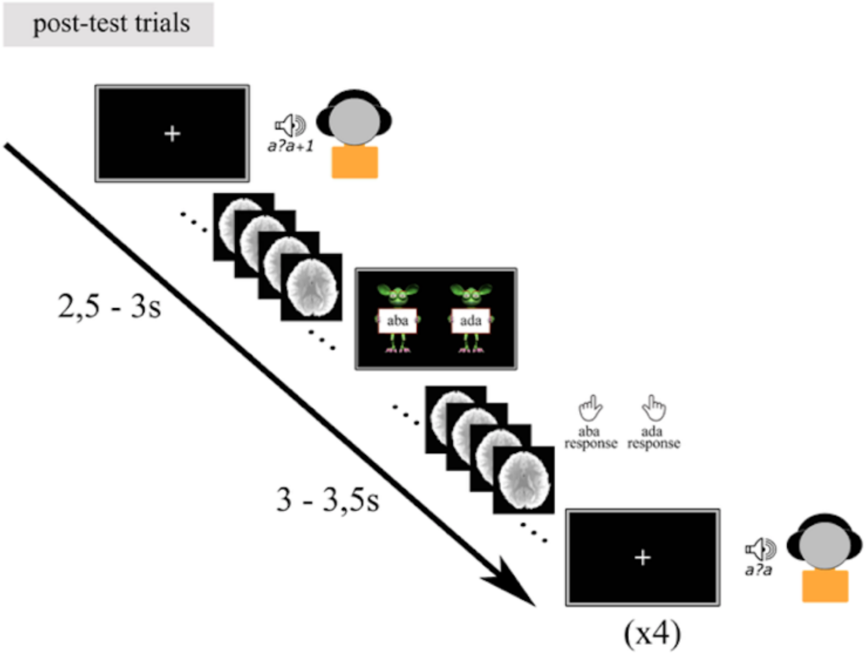


Note. This figure displays the experimental setup for the text-based (left) and lipread (right) audiovisual exposure blocks in the MRI scanner.

Post-test trials. In both, the behavioural and the scanning sessions, each exposure block was followed by four auditory-only post-test trials. Stimuli in the post-test trials were twice the most ambiguous sound /a?a/, as well as the flanking sounds /a?a/-1 (closer to /aba/) and /a?a/+1 (closer to /ada/), presented one time each. They were presented in randomized order and were followed by a visual response cue, identical to the one used in the pre-test (see figure 7). During the presentation of the sounds, participants saw a white fixation cross on a black screen. The onset of the response cue was jittered 2,5 s - 3 s following the presentation of the post-test sounds and had a fixed duration of 3 s. The subsequent post-test sound was presented 3 s - 3,5 s after the response cue. Thus, the total inter-trial interval between post-test sounds was 6 s.

Figure 7

Post-test trials



Note. This figure displays the experimental setup for the post-test trials in the MRI scanner.

Statistical analyses of behavioural data. Behavioural data were processed using *Python 3* (Van Rossum & Drake, 2009). In order to assess the performance of the participants in the recalibration task and subsequently the strength of the recalibration effect, we applied a within-subject design with condition, exposure, and post-test sounds as independent variables, and /aba/ response proportions as dependent variable. More precisely, we calculated a 2x2x3 repeated-measures ANOVA with the following within-subject factors:

- condition, i.e. type of visual stimulus during the exposure blocks (text-based vs. lipread)
- exposure, i.e. type of disambiguating sound during the exposure blocks (/aba/ vs. /ada/)
- post-test sounds, i.e. sound during the post-test trials (/a?a/ vs. /a?a/-1 vs. /a?a/+1)

Mauchly's test of sphericity was applied to calculate whether the assumption of sphericity was violated.

fMRI data acquisition

Brain imaging was performed using a Siemens Prisma 3T MRI scanner (Siemens Medical) using a 64-channel head-neck coil. Scanning time per participant was one hour. Besides the functional scans, a high-resolution structural scan (1mm x 1mm x 1mm resolution) using a T1-weighted three-dimensional MPRAGE sequence (repetition time [TR] = 2300 ms, echo time [TE] = 2.98 ms, 192 sagittal slices) was acquired. The recalibration task consisted of 5 runs, lasting approximately 7 minutes each, so that in total 5 functional scans (one per run) were acquired (2mm x 2mm x 2mm resolution) with a multi-band factor of 5 echoplanar-imaging (EPI) sequence ([TR] = 2000 ms, acquisition time [TA] = 1100 ms, field of view [FOV] = 210mm × 210mm, [TE] = 35.8 ms). Each volume consisted of 50 slices, covering the whole brain.

fMRI pre-processing. The software *BrainVoyager 22.0* (Brain Innovation, Maastricht, The Netherlands) was used for brain imaging data pre-processing and statistical analyses (Goebel et al., 2006). Inhomogeneity correction (with manual adjustments where necessary) was executed for the anatomical scans in order to obtain a clearer differentiation between white and grey matter. The anatomical scans were subsequently normalized according to the MNI template (Montreal Neurological Institute) in order to ensure comparability between subjects. For the functional data, 3D head motion correction with trilinear sinc interpolation was performed. For the motion correction, the first volume of the first functional run was used as reference slice in each participant respectively. Functional runs, in which the head motion exceeded a threshold of 3 mm rotation or translation in x, y, or z direction, were excluded from further analyses.

Subsequently, slice scan time correction (22 ms inter-slice time) and temporal high pass filtering (7 cycles per time course) were applied to the functional data. Slice scan time correction is necessary because the slices are not acquired sequentially and each slice group has a different delay leading to an overlap if not accounted for. The COPE 1.1 plugin (Breman et al., 2020) was used to correct for EPI distortions occurring due to non-zero off-resonance fields. COPE uses a sequence of images recorded with opposite phase-encoding direction than the rest of the recordings (in our case posterior to anterior instead of anterior to posterior) to estimate and correct distortions.

During co-registration, the distortion-corrected functional images were applied to the raw anatomical image. In this step, they were first aligned automatically, followed by manual fine alignment. Putting together the aligned files and the MNI-normalized anatomical image, normalized functional volume time course (VTC) files were created for each run. For this, they were re-sampled to 3 mm x 3 mm x 3 mm resolution.

Statistical analyses of fMRI data. We analysed the data obtained during the fMRI scans for neural activation during exposure blocks with condition as independent variable and changes in activation as dependent variable. A whole-brain analysis using a fixed-effects general-linear model (GLM) was applied to the data. By doing so, effects were averaged over all subjects. Even though insightful, this approach results in findings that cannot be generalized beyond the participants of our study. Functional contrasts in the GLM analysis were the following:

- text-based exposure blocks vs. baseline (fixation cross)
- lipread exposure blocks vs. baseline (fixation cross)
- text-based exposure blocks vs. lipread exposure blocks

Correlation analyses. We additionally looked at possible correlations between the neural activation pattern in selected regions of interest (ROIs) and the strength of the behavioural recalibration effect, as well as between the neural activation pattern with the scores from the Eén-Minuut-Test and the Spoonerism task. These exploratory analyses were run in BrainVoyager separately for the text-based and the lipread condition. The regions of interest were created using a built-in BrainVoyager function, which converts map clusters to voxels of interest (VOIs). The neural activation map used for the creation of our VOIs was the one displaying contrasts for lipread vs. text-based condition. For the lipread condition, these were ROIs in the bilateral visual cortex (peak coordinates in MNI space: x=1, y=-77, z=4), in the right motor cortex (peak coordinates: x= 48, y=-4, z=53), and in the left motor cortex (peak coordinates: x=-46, y=-8, z=52). For the text-based condition, the ROIs were located in the left visual cortex (peak coordinates: x=-28, y=-94,z=-6), the right visual cortex (peak coordinates: x= 24, y=-97, z=-8), and in the left frontal area (peak coordinates: x=-55, y=5, z=34).

Results

Behavioural recalibration task

A 2x2x3 repeated-measures (RM) ANOVA was used to calculate performance in the recalibration task. Within-subject factors were condition (lipread vs. text-based), exposure (/aba/ vs. /ada/), and post-test sounds (/a?a/ vs. /a?a/+1 vs. /a?a/-1).

The RM ANOVA revealed a significant main effect of exposure (aba vs. ada) [F(1,9) = 48.45, p < 0.001] and post-test sounds [F(2,18) = 56.27, p < 0.001], as well as a significant interaction effect between exposure and post-test sounds [F(2,18) = 7.88, p = 0.003]. There was no significant main effect for condition, and also no significant interaction effect between post-test sounds and condition or between exposure, post-test sounds, and condition. However, there was a significant interaction effect between exposure and condition [F(1,9) = 8.21, p < 0.05]. Table 2 displays F-values, numerator degrees of freedom (Num DF), denominator degrees of freedom (Den DF), and p-values (Pr>F) for each factor separately, as well as for the interactions.

Table 2

Results from the 2x2x3 repeated-measures ANOVA

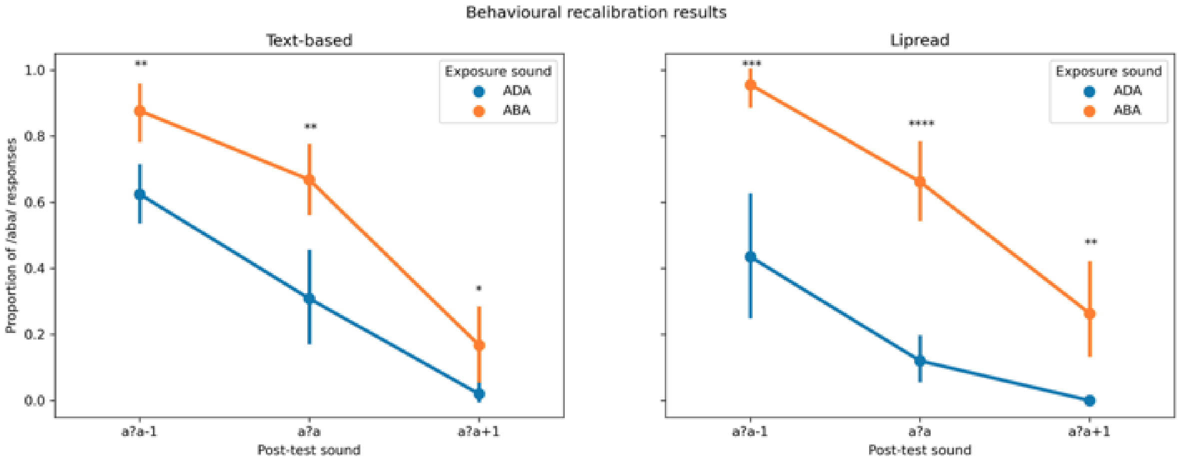
	F Value	Num DF	Den DF	Pr > F
Exposure (aba vs. ada)	48.4535	1.00	9.00	0.0001
Post-test sounds	56.2729	2.00	18.00	0.0000
Condition	1.3969	1.00	9.00	0.2675
Exposure (aba vs. ada) : Post-test sounds	7.8822	2.00	18.00	0.0035
Exposure (aba vs. ada) : Condition	8.2195	1.00	9.00	0.0186
Post-test sounds : Condition	3.0807	2.00	18.00	0.0707
Exposure (aba vs. ada) : Post-test sounds : Condition	0.3549	2.00	18.00	0.7061

Post-hoc paired sample t-tests were run on the proportions of /aba/ responses for each of the three post-test sounds following both exposure blocks, for each condition separately. The analyses showed a significant difference in the proportion of /aba/ responses following an /aba/ exposure block compared to an /ada/ exposure block across all three post-test sounds for the lipread condition (/a?a/: M = 0.66, SD = 0.18 vs. M = 0.12, SD = 0.11, t(9) = -7.02, p < 0.001; /a?a/+1: M = 0.26, SD = 0.23 vs. M = 0.0, SD = 0.0, t(9) = -3.41, p < 0.05; and /a?a/-1: M = 0.95, SD = 0.08 vs. M = 0.43, SD = 0.30, t(9) = -5.81, p < 0.001), as well as for the text-based condition (/a?a/: M = 0.66, SD = 0.16 vs. M = 0.30, SD = 0.24, t(9) = -4.14, p < 0.005; /a?a/+1: M = 0.16, SD = 0.18 vs. M = 0.02, SD = 0.04, t(9) = -2.50, p < 0.05; and /a?a/-1: M = 0.87, SD = 0.13 vs. M = 0.57, SD = 0.19, t(9) = -3.73, p < 0.005).

The significant main effects of exposure (aba vs. ada) and post-test sounds, as well as their interaction indicate that the paradigm elicited a significant recalibration effect. The significant interaction between exposure (aba vs. ada) and condition indicates that the response proportions varied according to whether the disambiguating visual information (/aba/ or /ada/) was presented in form of text or lip movements. However, given the lack of a threefold interaction of exposure (aba vs. ada), post-test sounds and condition, these findings do not hint towards significant differences in recalibration effect when comparing the lipread and the text-based condition. Nevertheless, visual inspection (see figure 8), reveals a slightly stronger recalibration effect for the lipread condition since the difference of /aba/ response proportions following /aba/ vs. the ones following /ada/ exposure is bigger.

Figure 8

Results of the behavioural recalibration task



Note. This figure displays the results of the behavioural recalibration task in the text-based (left) and the lipread (right) and condition. The y-axis indicates the average proportion for /aba/ responses to the three post-test sounds displayed on the x-axis after "aba" exposure blocks (orange) and "ada" exposure blocks (blue). Vertical bars represent the standard error (* = p<=0.05, ** = p<=0.01, *** = p<=0.001, **** = p<=0.0001).

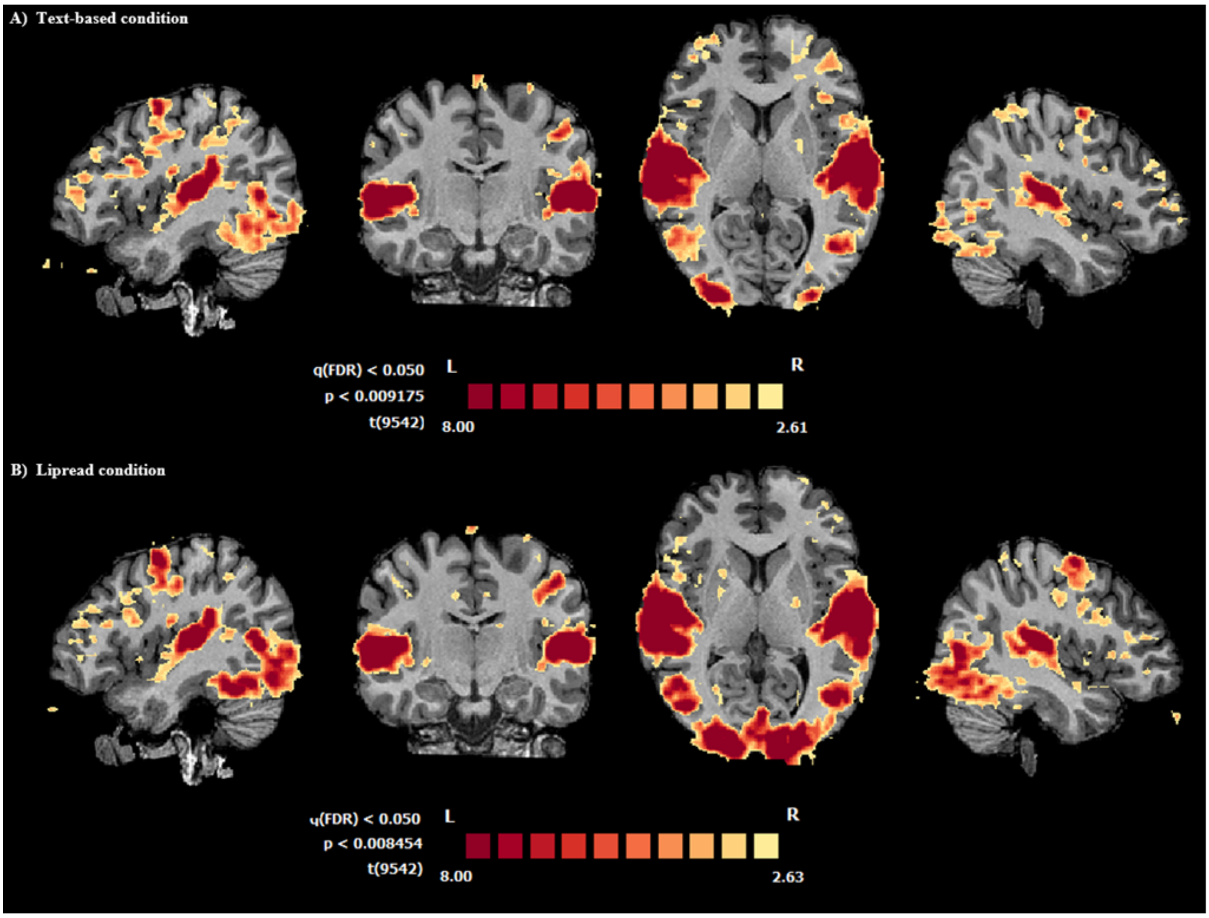
Boxplot-analyses revealed an outlier for the behavioural results of the lipread recalibration task. This participant showed a significantly higher recalibration effect than the other subjects and was therefore excluded in the analysis of the behavioural data. However, excluding this outlier from the fMRI analyses did not significantly change the neural activation pattern, so that the data of the participant was not excluded there.

Neural activity during exposure blocks

A fixed-effects (FFX) general linear model (GLM) analysis was applied in order to determine cortical activation during the exposure blocks. There was a broad activation pattern in regions associated with reading and audio-visual integration for both, the text-based (see figure 9A) and the lipread exposure condition (see figure 9B).

Figure 9

Results of the fMRI analysis for both conditions separately vs. baseline

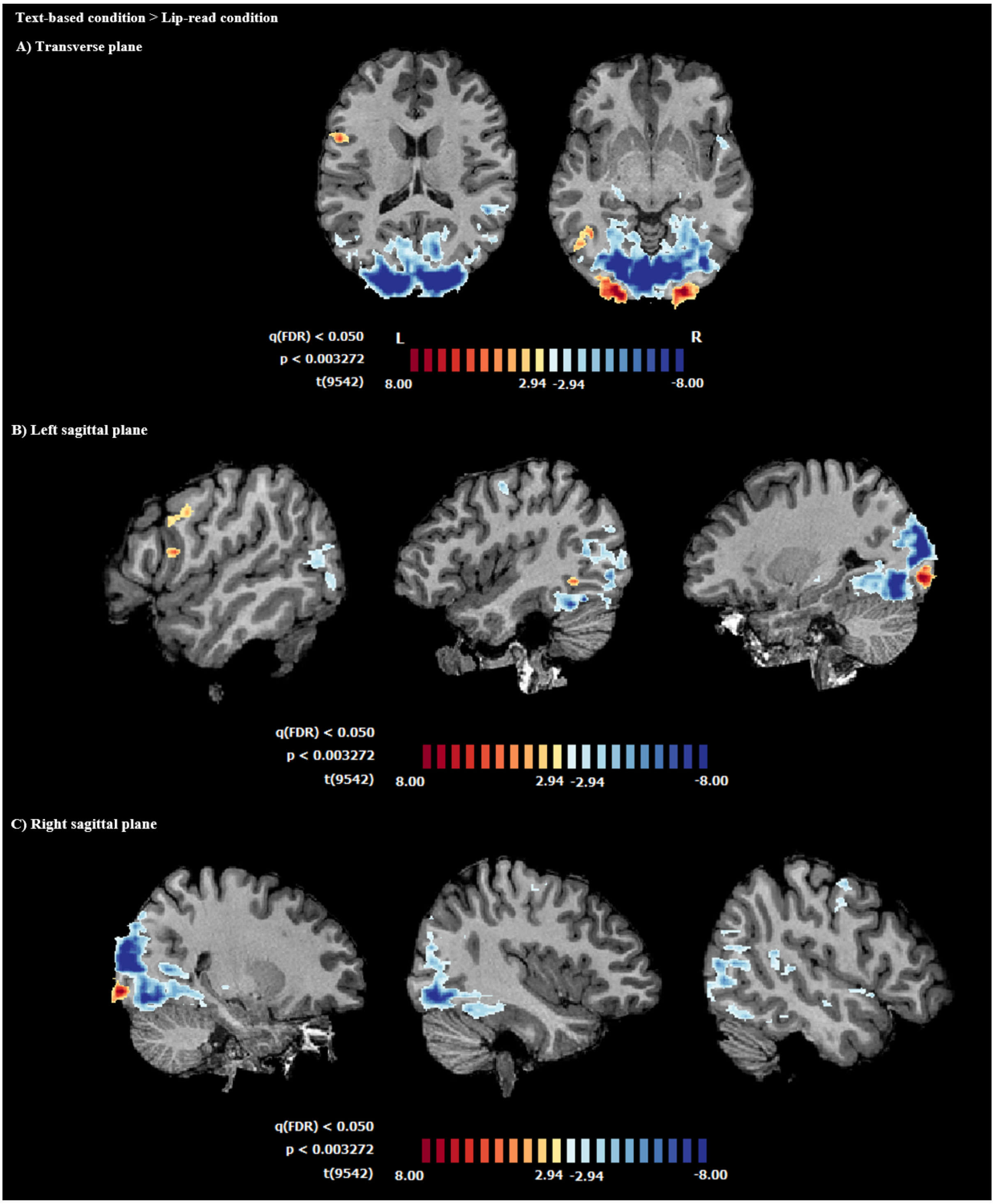


Note. This figure displays the brain activation pattern during (a) text-based exposure block vs. baseline; (b) lipread exposure block vs. baseline (L = left, R = right).

The activated regions included the occipital (visual) cortex and the fusiform area, the primary auditory cortex with bilateral superior temporal gyrus (STG) and sulcus (STS), as well as parietal areas and the premotor, supplementary motor and visual motor cortex. Even though both conditions elicited broad bilateral activation, visual inspection further revealed a tendency towards stronger activation in the left hemisphere for the text-based condition.

Figure 10

Results of the fMRI analysis for lipread vs. text-based condition



Note. This figure displays the brain activation pattern for lipread exposure block (blue) vs. text-based exposure block (red) in (a) transverse plane; (b) left sagittal plane; (c) right sagittal plane (L = left, R = right).

The main difference between conditions, was a stronger activation of the bilateral occipito-temporal cortex for the lipread condition (see figure 10A). Further, the activation cluster in the occipital cortex was more frontal than in the text-based condition including the left and right fusiform areas. The activation in the text-based condition was stronger in the posterior occipito-temporal cortex including visual area 1 (V1) and there was also activation in the left fusiform area. Besides that, there was a small cluster in the left inferior frontal gyrus (IFG) that showed a higher activation in the text-based condition compared to the lipread condition (see figures 10A and 10B). There was also activation for both conditions in motor areas, with the text-based condition eliciting activity in more lateral left pre-motor and supplementary motor areas (see figure 10B), while the lipread condition elicited bilateral activation in more superior areas (see figures 10B and 10C).

Correlation analyses

Exploratory correlation analyses were run in BrainVoyager separately for the text-based and the lipread condition, in order to investigate possible correlations between the strength of the behavioural recalibration effect and activity in selected regions of interest. Additionally, we ran correlation analyses for the same ROIs with the scores from the Eén-Minuut-Test and the Spoonerism task. These ROIs were located in the bilateral visual cortex and in the bilateral motor cortex for the lipread condition, and in the bilateral visual cortex and the left frontal area for the text-based condition. However, the analyses revealed no significant correlation for the strength of the recalibration effect with activation in any of the regions of interest, neither for the lipread nor for the text-based condition. Similarly, correlations between cortical activation in these regions of interest and participant's scores on the Eén-Minuut-Test and the Spoonerism task were not significant.

Discussion

The present study investigated neural correlates of lipread and text-based recalibration in dyslexic adults. There was a significant recalibration effect in both, the lipread and the text-based condition. The strength of this effect did not differ between conditions. The fMRI analysis showed a comparable brain activation pattern during exposure blocks in both conditions. However, there were also differences in activation in the occipital cortex, where the lipread condition elicited a broader activation pattern including the fusiform face area, while the text-based condition was associated with activity in more posterior areas, as well as in the visual word form area. Besides that, we found differences between the conditions in activation of the premotor cortex. Further, only the text-based condition elicited activation in the left inferior frontal gyrus.

Behavioural recalibration effect

Significant text-based and lipread recalibration effect. Behavioural results indicated both, a text-based and a lipread recalibration effect in dyslexic adults. This confirms the findings by Baart et al. (2012), who found a lipread recalibration effect in dyslexic adults, and Romanovska et al. (2019, 2021), who reported a text-based recalibration effect in dyslexic children. Previously, Keetels et al. (2018) did not find a text-based recalibration effect for dyslexic readers. This discrepancy in findings might be due to a difference in paradigm design: Keetels et al. (2018) combined exposure blocks containing ambiguous and exposure blocks containing unambiguous information, while in the research done by Romanovska et al. (2019), they were kept separate. Combining ambiguous sounds with unambiguous visual information elicits recalibration effects, while using an unambiguous speech sound in combination with an unambiguous text stimulus leads to a so-called selective adaptation effect or repulsive bias. This describes a shift in perception in the opposite direction Vroomen et al. (2007). In the case of our paradigm, repeated exposure to /aba/ speech sounds in combination with "aba" text would lead to a tendency to perceive /ada/. It has been found that the type of exposure has different influence on the recalibration and adaptation effects, so that the adaptation effect becomes stronger with more exposure while the recalibration effect shows a curvilinear behaviour, being initially strong and then weakening until it disappears (Vroomen et al., 2007). In our study, we did only use ambiguous speech sounds, focusing exclusively on recalibration effects. Keetels et al. (2018) in their study, on one hand mixed combinations of exposure blocks and post-test trials leading to recalibration and those leading to adaptation, and on the other hand they also mixed text-based and lipread exposure trials while we presented them in blocked manner. This might explain the discrepancy between our findings and the ones obtained by Keetels et al. (2018).

Similar strength of text-based and lipread recalibration effect. Unlike predicted, we did not find a significant difference between the strength of the recalibration effect between conditions. We initially hypothesized the effect to be stronger for the lipread condition (cf. Keetels et al., 2018; Ullas, Formisano, et al., 2020) assuming that impaired phoneme-grapheme conversion in dyslexic readers would lead to worse performance when text serves as disambiguating information. Even though not significant, visual inspection indicated a marginally stronger effect for the lipread condition, which goes in line with our hypotheses.

The fact that there was no significant difference between conditions might again be caused by a choice in experimental design. In each run, we presented the four text-based and the four lipread exposure blocks together and not in alternating or scrambled manner. Using a similar paradigm, Keetels et al. (2018) had previously reported no text-based but a lipread recalibration effect for dyslexic readers. In their study, they varied the audiovisual exposure condition every ten items. This might explain the divergent findings, since in our study participants were presented with the same audiovisual exposure condition for a duration of 24 trials, interrupted by post-test trials after every 8 trials. Therefore, the difference between conditions might not have been as pronounced in our study compared to Keetels et al. (2018), because participants had enough time to get used to a certain type of presentation of the disambiguating stimulus.

Another reason for finding comparable recalibration effects in the text-based and the lipread condition, could simply be that phoneme-grapheme conversion as measured by the recalibration task, is not impaired to an extent in dyslexic readers, that would lead to worse text-based recalibration effects. It might be another aspect of phonetic processing that is impaired more crucially leading to reading difficulties in dyslexic individuals.

On the other hand, it could also be that dyslexic individuals have difficulties with both, lipread and text-based recalibration and perform lower than non-dyslexic readers in both tasks, which would hint towards a more general auditory-visual integration deficit, and not only a specific one for orthographic information. However, as we were able to detect a significant recalibration effect for the dyslexic group but lack a reference point, only a direct comparison of dyslexic and non-dyslexic readers can shed light on this dubiety.

Potentially stronger lipread recalibration effect. Visual inspection of the results revealed a marginally stronger recalibration effect in the lipread condition. Previous studies explained the stronger lipread recalibration effect by the developmentally older connection between lip movements and sounds as opposed to the learned arbitrary correspondence between text and speech sounds (Keetels et al., 2016). On the other hand, text contains more clear phonetic information than lip movements, which are prone to be more ambiguous. For example bilabial closure can correspond to either the sounds /b/, /p/ or /m/ while the grapheme "b" exclusively corresponds to /b/ (Keetels et al., 2016). Our results suggest that in dyslexic readers, these two modalities of association seem to be similarly processed and automatized, and lead to comparable perceptual shifts in the phonetic recalibration paradigm. The fact that we found a trend towards a stronger lipread recalibration effect, goes in line with previous findings (cf. Keetels et al., 2018). Yet, it needs to be kept in mind that in our study, the difference in effect strength between conditions was not significant and should therefore not be over-interpreted.

Text-based recalibration effect in adults. So far, a text-based recalibration effect was only found dyslexic children (cf. Romanovska et al., 2019, 2021). This was explained by the flexibility of cortical systems involved in letter-speech sound integration, which - while still in development - are hypothesized to be more flexible in the perception of phonemic categories, both in dyslexic and typically reading individuals. In our study, we found text-based phonetic recalibration to be present also in dyslexic adults. This shows that at least in this type of paradigm, phoneme-grapheme conversion does not seem to be impaired to an extent that would hinder shifts in the perception of ambiguous sounds. One possible explanation for this finding is that dyslexic adults developed sufficient compensation strategies during their lifespan in order to account for difficulties during phoneme-grapheme conversion. However, such compensation strategies would most likely rely on context information or morphological and semantic aspects of the processed words (Rasamimanana et al., 2020), which in the case of our study, where non-words were used as stimuli, was arguably not possible.

Choice of stimuli. Further, the design of the stimuli might have had an influence on our results. Simple two-syllabic non-words were chosen on purpose as stimuli since no lexical retrieval takes place during their processing. By doing so, artifacts due to semantic and lexical processing could be ruled out. Our study indicates that the phoneme-grapheme conversion of simple syllables does not seem to be impaired in dyslexic readers in a manner that would prevent perceptual shifts to occur. However, this might still be the case when dyslexic readers are presented with real words, shorter, or longer chunks of text.

In a study with adult dyslexic readers Blau et al. (2009) and dyslexic children (Blau et al., 2010), the authors used single letters and three-letter combinations as stimuli in combination with matching and non-matching sounds. fMRI results revealed hypo-activation of the superior temporal cortex during this task, which required the integration of text and speech sounds of single letters. Further, reduced audio-visual integration correlated significantly with lower performance in phonological tasks. The authors conclude therefore that impairments in audiovisual integration in dyslexic readers can be explained by more fundamental auditory processing deficits. Using ERP and a similar paradigm as Blau et al. (2009), Žarić et al. (2014) were also able to show that reduced neural integration of letters and speech sounds significantly correlates with reading abilities in dyslexic children. Kronschnabel et al. (2014) investigated audiovisual integration in dyslexic adults and found that the more naturalistic and word-like the stimuli, the more pronounced were processing deficits in their study.

Given those findings, our paradigm might therefore not have been tailored adequately to detect the whole range of impairments regarding phoneme-grapheme conversion in dyslexic readers, and further studies could do so by for example comparing non-words to real words using a similar paradigm.

Neural activation

As hypothesized, we found a similar neural activation pattern during the lipread and the text-based exposure blocks. Activation was found in a broad bilateral network associated with reading and audiovisual integration (cf. Kilian-Hütten et al., 2011; Ullas, Hausfeld, et al., 2020). More precisely, neural activity was found in the occipital cortex, the superior temporal cortex, the parietal cortex, and motor areas. This cortical activation pattern goes in line with findings from previous fMRI studies using a similar phonetic recalibration paradigm (Bonte et al., 2017; Romanovska et al., 2021). Further, results overlap with findings from studies investigating letter-speech integration in adults (cf. Blau et al., 2009). In general, our results indicate that the cortical networks recruited during exposure to ambiguous speech in combination with text are mostly similar to those involved in the processing of ambiguous speech combined with lip movements. However, there were some differences in activation between the two conditions in the occipital cortex, the premotor cortex, and inferior frontal areas. In the following, the activated brain areas and their associated functions will be described in more detail, and by doing so, their functional role in our study will be assessed.

Areas activated in both conditions.

In both conditions, we found broad activation in the bilateral occipital cortex. Given the visual nature of our stimuli in both conditions and results from previous studies (e.g. Kilian-Hütten et al., 2011), we had hypothesized the occipital cortex, the primary centre of vision in the human brain, to be active.

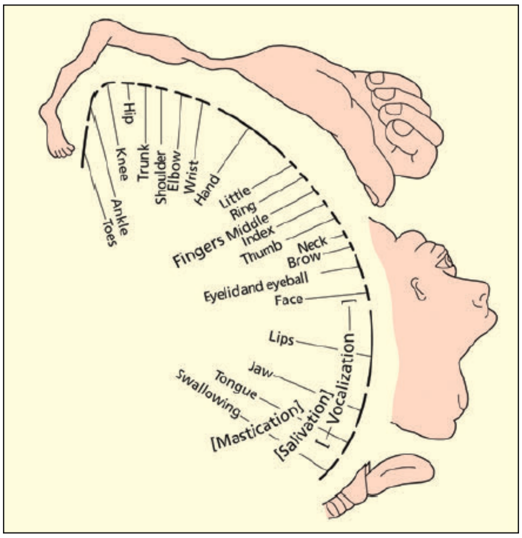
Furthermore, the superior temporal cortex was activated bilaterally in both conditions. This area, commonly referred to as auditory cortex, includes the superior temporal sulcus (STS) and gyrus (STG), as well as the planum temporale (PT). We expected to find activation in this area since our stimuli in both conditions consisted not only of a visual, but also an auditory input component. The STS is part of the dorsal temporo-parietal reading network and has been found to be involved in multimodal integration. This was for example shown in a study by Van Atteveldt et al. (2007), where the STS showed increased activation in response to bimodal stimuli as opposed to unimodal ones. Given its key role in information integration, the STS is naturally also involved in the processing of audiovisual information, such as the mapping of phonemes and graphemes (Raij et al., 2000). In our study as well, participants were required to simultaneously integrate information from two different modalities (vision and audition). We expected activation in the STS and parietal areas in general, and were indeed able to detect it in both conditions. Blau et al. (2009) investigated the integration of letters and speech sounds comparing dyslexic to typical readers, and found that even though auditory areas are still activated for both participants groups, dyslexic readers tend to under-activate the superior temporal cortex. Since in our study, all participants were dyslexic, we cannot indicate whether the activation was weaker than it would have been for typical readers. However, activation in the auditory cortex does not seem to differ across the type of audiovisual integration condition.

Finally, in line with previous studies (e.g. Ullas, Hausfeld, et al., 2020), we also expected to find activation in motor areas. Even though at first glance, the involvement of the motor cortex seems counter-intuitive for a passive exposure task, there are several explanations for activation in these areas. It is a well-known finding that the visual perception of actions is associated with activation in motor areas even though no action is actually executed by the participant (Watkins et al., 2003). This activation in the so-called mirror neuron system is hypothesized to serve the purpose of facilitating recognition, understanding, and imitation of actions. In our study, especially the video condition might have elicited this type of activation. However, it has also previously been found that premotor and motor areas, as well as parieto-temporal areas are active during speech processing. In that case, they serve the purpose of mentally representing articulatory gestures and sensorimotor functions (Meister et al., 2007). This is based on the theory that speech initially evolved from gestural communication, which is represented and processed in form of observation-execution matches in the mirror neuron system (Murakami et al., 2011). The articulatory system is therefore active during speech sound mapping in a way that participants sub-vocally repeat a word or sound they are exposed to, leading to activation in motor areas. Especially the dorsal auditory pathway including sensorimotor and somatosensory regions has been found to be active during this process (Correia et al., 2015). Watkins et al. (2003) found both, auditory and visual speech perception to be associated with activity in parts of the motor system generally involved with speech production. In our study, there was indeed activation in motor areas for both conditions, underlining the involvement of these areas in auditory speech processing and speech sound mapping. However, given the audiovisual nature of our stimuli, it is not fully traceable whether this activation was elicited more by the visual or more by the auditory component of the stimulus, or the integration of the two. However, Kilian-Hütten et al. (2017) argue in favour of multisensory rather than unimodal processing mechanisms being in place during audiovisual processing, so that a further disentanglement of the exact reason for neural activation does not seem necessary.

Another - yet unlikely - explanation for the activation of motor areas, is that it is an artifact due to accidental button presses during exposure blocks. Even though participants were clearly instructed to only passively view and listen to the stimuli, they might have been mislead since they were required to press a button during the post-test trials. Since after eight exposure trials, there was always a block of four post-test trials presented, participants might have been inclined to press a button also during the exposure block trials leading to activation in the motor cortex. The primary motor cortex is divided by functions corresponding to different body parts in a somatotopic manner, including hands and facial areas (see figure 11). The fact that the exact location of the active motor regions is closer to the so-called hand knob or hand motor hot spot in the primary motor cortex, and not overlapping with the slightly more anterior mouth region, which is the area where we might expect more activation in sub-vocal speech sound production, would speak in favour of this explanation. An inspection of every participant's responses during the entire scanning session revealed however, that there was no above-average amount of erroneous button presses. Taking together all participants and all runs, there were only sixteen button presses during exposure blocks. Nine of those were executed by the same participant at the very beginning of the scanning session, and can thus be explained by initial insecurities regarding the correct time for the button press. Taken together, this indicates that activation in the motor cortex is not ascribable to erroneous button presses during exposure blocks.

Figure 11

Somatotopic division of the primary motor cortex



Note. This figure (taken from Ward, 2019) displays the primary motor cortex with the respective body parts controlled by a certain area.

Differences in activation between conditions. One of the main differences between conditions, was the activation pattern in the occipital cortex. While the lipread condition elicited broad activation in the bilateral occipito-temporal cortex, the text-based condition led to significantly higher activation bilaterally in the posterior part of the occipital cortex, as well as in the left fusiform area. These differences can be explained through the nature of the visual component of the stimuli in the two conditions.

The video stimulus in the lipread condition contained rich and salient visual information like faces and movements. Compared to text, this more complex stimulus is prone to lead to higher attentional capture and subsequent stronger overall activation of the visual cortex and adjacent temporal areas. The activated areas further extend bilaterally to the fusiform gyrus, which is located in the inferior temporal cortex and seat to the fusiform face area (FFA). The FFA area has been proposed by Kanwisher et al. (1997) to be selectively active when participants are exposed to faces as opposed to other types of visual information. According to Grill-Spector et al. (2004), the center of the FFA is located at the following position in the brain: MNI coordinates in x, y, and z direction: right hemisphere: 41 ±3, -39 ±7, -24 ±5; left hemisphere: -40 ±4, -41 ±7, -23 ±5. In our study, we found activation in the area of the right fusiform area at the following MNI coordinates: 41 -41 -21. In the left hemisphere, we found activation at the following MNI coordinates: -40 -46 -23. Thus, we did indeed find bilateral activation in areas almost precisely overlapping with region referred to as fusiform face area. Activation of the FFA during the lipread but not during the text-based condition in our study, therefore shows the specific involvement of this area in the processing of faces, which is in line with the literature.

The text-based condition on the other hand, elicited activity in more posterior occipito-temporal regions, which include the primary visual cortex (V1). The V1 is involved in the processing of more basic visual information such as edges and bars (Seydell-Greenwald et al., 2021). The stimuli in the text-based condition consisted of strings of letters, which in substance are configurations of lines. Stronger activation of the V1 during the text-based condition as compared to the lipread one, can thus be explained by its involvement in the processing of written text. In a recent study, Seydell-Greenwald et al. (2021) found the primary visual cortex to also be active during the processing of spoken language, most likely serving the function of predicting possible subsequent visual input. Due to the bimodal nature of our stimuli, it is again difficult to exactly derive whether activation in a certain area stems from the auditory or the visual input or their integration. It is therefore possible that the V1 was not only activated by the text stimulus, but also by the speech sound. However, this does not explain the significantly higher activation during the text-based condition since the same auditory stimulus was present in both conditions. Taken together, this makes a specific influence of written text on the activation of the primary visual cortex likely.

An area that has been found to be selectively involved in the processing of written text, is the visual word form area (VWFA). The VWFA is located in the left fusiform gyrus and is part of the ventral occipito-temporal reading network. Interestingly, the VWFA shows comparable activation for both, real words and readable non-words, which seems to be necessary for reading unknown words (Cohen et al., 2002). Previous studies (cf. Cohen et al., 2002) located the peak of the VWFA at the following MNI coordinates in x, y, and z direction: –46 –53 –20, with a standard deviation of approximately 0.5 cm. In our study, we found activation at the following MNI coordinates: -46 -53 -8. This area is in the range of the area defined as VWFA, which indicates a selective involvement in the processing of text stimuli as opposed to video stimuli and is again in line with previous literature.

However, it should be mentioned that the existence of both, the VWFA and the FFA, have been subject to debate since their proposal in the scientific community. Regarding the visual word form area, Price and Devlin (2003) argued that the area is also active during tasks that do not involve word processing. Further, people suffering from pure alexia, who are not able to read but whose writing skills are not impaired, have not been found to have specific damage limited only to the VWFA. The authors further argue that rather than limiting the processing of words to a small specialized area in the brain, it would be worthwhile to investigate the role of the VWFA in a broader word processing network including several interacting regions. For the case of the fusiform face area, Weiner and Grill-Spector (2012) argue in a likewise manner that there supposedly is a broader network of face-selective regions in the ventral temporal cortex and that there is a large variability in the localization of the FFA. Again, it seems that rather than attributing a very specific function to a certain brain area, it is more valuable to investigate interactions in the sense of a broader cortical network involved in the processing of faces.

Additionally, only for the text-based condition, we found activation in the left inferior frontal gyrus (IFG). The left IFG is part of the ventral occipito-temporal reading network, which is involved in more automated and fast processing of words. Further, the left IFG is associated with speech comprehension and with unifying different levels of linguistic information (Ullas, Hausfeld, et al., 2020). The left IFG has been found to be active together with the STG when people experience the McGurk effect, which again indicates an involvement of the left IFG during the attempt to comprehend speech and the use of additional contextual information (Ullas, Hausfeld, et al., 2020). Romanski (2012) found the left IFG to be involved in the integration of audiovisual information in form of facial gestures and speech. This seems to hold true not only for bodily but also for textual information, given that the left IFG has been found to play a role in phoneme-grapheme conversion and phoneme assembly during reading (Joubert et al., 2004). In our study, the text-based condition elicited significantly stronger activation in the left IFG than the lipread condition. Therefore, it can be assumed that even though its involvement in speech comprehension, and more precisely the integration of audiovisual information, the left IFG seems to be particularly engaged during the association of phonemes and graphemes, as it was necessary during the text-based condition. Another possible explanation for the activation of the left IFG, is its role in compensation mechanisms in individuals with reading impairments. Together with the right IFG and the homolog of the VWFA in the right hemisphere, the left IFG has been found to be overactive in dyslexic readers, while the left hemisphere posterior reading system is impaired (Shaywitz & Shaywitz, 2008). Activation in the left IFG for the text-based condition in our study could therefore also be due to the presence of compensatory mechanisms during the processing of text and its associations with speech sounds. The employment of compensation strategies would not be necessary when observing lip movements and integrating them with speech sounds, since there is no impairment for this type of audiovisual integration in the first place. However, since the VWFA is usually underactivated in individuals with dyslexia, the significant involvement of this area during processing of the text-based condition, might be better explained with the nature of the stimulus rather than its role for compensatory mechanisms. Finally, we also found significant differences in activation in motor areas. For the text-based condition, activation clusters were placed more laterally and only in the left hemisphere, while activated motor areas in the lipread condition were bilateral and at a more superior position close to the supplementary motor area. The premotor cortex can be subdivided into four smaller, more specialized areas, which have been investigated mainly in animal studies. The subdivision of these areas follows a logic according to their exact location in the premotor cortex, that is depending on whether they are located more dorsal or ventral, as well as more rostral or caudal. In our study, the activation cluster in the text-based condition overlaps broadly with the premotor dorsal rostral area (PMDr). The PMDr has been found to be involved in learning arbitrary associations of sensory stimuli with specific movements (Brasted & Wise, 2004). The areas with significant activation for the lipread condition, were located roughly in the area of the premotor dorsal caudal cortex (PMDc), and extend further towards the supplementary motor area. In a direct comparison on the involvement of these areas on verbal processing using fMRI, the PMDr has been found to be more active during a non-linguistic mental operation task as compared to verbal rehearsal, while the PMDc was equally active in both conditions (Hanakawa et al., 2003). The authors conclude that the PMDr is more involved in cognitive mapping, while the PMDc is associated more directly with motor control and less with associated cognitive functions. The fact that the PMDr was more active in the text-based condition in our study, would speak in favour of this account since the presentation of text by nature requires more abstract processing than lip movements. The supplementary motor area is associated with movement control, which could explain its activation in the lipread condition, where participants are required to execute stronger inhibition in order not to mimic the lip movements they view on the screen.

In general, we found a more or less evenly distributed bilateral activation pattern for both conditions. However, the text-based condition elicited slightly more pronounced activation in the left hemisphere, for example in the case of the inferior frontal activation. This finding goes in line with the commonly supported account that language and associated functions such as reading are left-lateralized in the brain. More precisely, Hickok and Poeppel (2007) found the ventral stream to be bilateralized during speech processing, while the dorsal stream seems to more dominant on the left hemisphere. Given that both conditions involved language processing, it seems that especially the processing of text as opposed to lip movements seems to elicit more activation in the left hemisphere. Previous studies found a greater involvement of the right hemisphere in dyslexic readers, which together with activation in frontal areas, is thought to be due to development of compensatory mechanisms (Sandak et al., 2004). While we found activation in inferior frontal areas, we did not find a higher activation in the right hemisphere. Therefore, we are unable to draw conclusions about the connection between lateralized activation and possible compensatory mechanisms during speech processing.

As previously established, the reading network comprises two posterior pathways: a dorsal (temporo-parietal) and ventral (occipito-temporal) one. Additionally, the inferior frontal gyrus, located slightly more anterior, constitutes a separate system (Sandak et al., 2004). In our study, activation for the text-based condition was found in the VWFA, which is part of the ventral reading network employed in automated text processing and in the left IFG, which is part of the anterior network employed during phonological recoding. During reading development, children first rely more on dorsal and anterior areas, while engagement of the ventral system is associated with increasingly skilled reading (Sandak et al., 2004). In dyslexic readers, the dorsal reading pathway has been found to be more active since it is involved in the systematic assessment of phoneme-grapheme connections and less automatized reading processes (cf. Shaywitz & Shaywitz, 2008). One could argue that our task design is similar to early reading mechanisms since the connections between the ambiguous sound and the disambiguating text information are novel and not yet established. This is similar to reading development, where connections between sounds and text are initially established and subsequently automatized. Interestingly however, in our study the ventral and anterior networks seem to have been more active than the dorsal one. Nonetheless, it could be possible that rather than a stronger employment of the entire ventral reading network, activation in the VWFA could be due to text processing per se and activation in the left IFG could be due to compensation mechanisms.

Given the similarity between the stimuli in the two conditions and the underlying processes involving audiovisual integration, we hypothesized to find a similar activation pattern for the two conditions in our study, which was indeed the case. However, it was one of our goals to also explore possible differences between the brain activation in the two conditions, in order to potentially be able to draw conclusions about the different nature of integration of text and speech vs. the integration of lip movements and speech in dyslexic readers. This would help to shed a light on the question whether dyslexia is associated with general audiovisual integration deficits or rather more specific ones involving phoneme-grapheme associations in particular. However, the differences we found in activation between the two conditions, were mostly due to the nature of the stimuli themselves. The similarities in neural activation, which go in line with the comparable behavioural effects, can either be explained by a similarly strong impairment in audiovisual integration beyond phoneme-grapheme conversion, including also the processing of lip movements in combination with speech. If that was the case, our participants would simply have performed equally low in both conditions, which due to the lack of a non-dyslexic control group, is hard to determine. However, we did find a significant correlation effect for both conditions, which indicates that our participants have sufficient audio-visual integration capacities for phonetic recalibration to occur.

Alternatively, it might be possible that our participants do indeed show impaired phoneme-grapheme conversion but during their lifespan were able to develop sufficient compensation strategies for there not to be a difference in the behavioural recalibration task. The areas activated during the text-based condition, speak in favour of the employment of such compensation mechanisms.

The absence of correlation effects. Unlike predicted, we did not find a correlation between the strength of the behavioural recalibration effect and neural activation in selected regions of interest that we chose because they had previously been found to show activation differences between the two conditions. In line with findings by Ullas, Hausfeld, et al. (2020), we had initially assumed to find a correlation between brain activation and recalibration task performance in a way that stronger neural activation would be associated with stronger recalibration effects. In their study, Ullas, Hausfeld, et al. (2020) had found this correlation only for the stronger lipread recalibration effect, but not for the weaker lexical one. However, their lexical condition differed from our text-based condition in a way that processing relied more on lexical and semantic knowledge due to the use of real words and pseudowords as stimuli. They explained this discrepancy in correlation effects by the generally lower performance in the lexical recalibration task, which lowered the scope for a significant difference between high and low scores within this condition, subsequently leading to an absence of correlation with neural activation. In our case however, the underlying processing mechanisms between the conditions do not differ that much since text-based recalibration is closer to lipread recalibration than the lexical recalibration used by Ullas, Hausfeld, et al. (2020). It could however still be the case that in our study there was generally a low recalibration effect in both conditions, not strong enough for our tests to be sufficiently sensitive in order to detect correlations with neural activation. This assumption can only be put into context by comparing the recalibration effect in our sample to the one of non-dyslexic participants, and see whether the latter group shows larger recalibration effects correlating with neural activation patterns. Importantly, also the small sample size most certainly had an influence on the absence of correlation effects in our study. It is possible that inter-individual differences in the strength of the recalibration effect were not large enough irrespective of a strong or weak effect on the group level, resulting in an impossibility to detect correlations with neural activation in selected ROIs. Furthermore, we did also not find a significant correlation between the results from reading assessment tests and neural activation patterns for both conditions separately, which shows that there was no direct connection between reading skills and neural activation during text-based or lipread exposure blocks. This again, might have been due to small number or participants and subsequent low statistical power, or the relatively uncomplex way of assessing reading skills. Another explanation could be that there was relatively large variation in participant's performance in the reading tasks. This is especially the case for the Spoonerism task, where values ranged between 65 and 212.7 with a mean of 121.83 and a standard deviation of 44.94. At least for this sub-task, the large variation could explain the absence of correlation effects.

Limitations and further directions

One of the main limitations of our study is its small sample size (n=10). The reason for this is that the present study is part of a larger project, where the performance and neural activation of dyslexic readers during the aforementioned recalibration task is compared to typical reading adults. This study itself is a pilot study involving only 20 participants. Since we exclusively focused on dyslexic readers in our study, we only included half of the sample. Due to the small number of participants, we could not apply a random-effects analysis on the fMRI data but had to use a fixed-effects analysis. This type of analysis does not allow for generalizations of the findings onto the general population. Our findings are therefore restricted to the sample under investigation, which represents a limitation in terms of generalizability of our results. Also on statistical analyses applied in order to obtain behavioural results, a small sample size has considerable influence. The fact that we did not find a significant difference in strength of recalibration effect between conditions was due to an absence of an interaction effect between exposure, post-test sounds, and condition in the ANOVA. In order to produce interaction effects of the same size as main effects, the sample size should be inflated fourfold and for interaction effects half the size of the main effect, the sample size would need to be approximately 16 times bigger (Brookes et al., 2004). The reason for not detecting an interaction effect and therefore no difference between conditions could therefore be due to a lack of statistical power. Accordingly, this holds true also for the absence of a correlation between behavioural and neural results. In future studies it would therefore be desirable to apply the same paradigm on a larger sample and ideally also compare dyslexic to non-dyslexic readers.

Another very crucial point regards the question whether the applied recalibration tasks is actually measuring what we would like to measure. It is possible that besides (or instead of) audiovisual integration processes, which lead to phonetic recalibration, our task paradigm rather measured the automatization of perceptual learning. We would thus not be able to trace back the influence of different mechanisms of audiovisual integration on phonetic recalibration, but rather just how well participants are able to make connections between two different types of input during the exposure blocks and use these skills to succeed in the post-test trials.

Unlike in many studies in the field of dyslexia, which focus on children and reading development, our subjects were dyslexic adults. It is to be assumed that dyslexic adults, especially the ones who participate in psychological research and circulate in the academic field, already acquired a certain level of literacy. This goes hand in hand with the development of considerable compensation strategies for their reading impairments, partially because they were simply exposed to larger amounts of text throughout their life. It would therefore be worthwhile to search for dyslexic adults with a lower level of formal education and compare their performance in a future study. Furthermore, it would be interesting to investigate the development of possible compensation strategies in readers with dyslexia throughout the lifespan. A longitudinal study on dyslexic individuals investigating changes in reading performance combined with the recalibration paradigm that we employed in our study and that Romanovska et al. (2021) successfully used in children, would therefore be valuable.

Our participants were native Dutch speakers. Dutch is an orthographically transparent language, meaning that graphemes map directly to a corresponding phoneme, which is supposed to facilitate reading. Thus, it would be interesting to see whether speakers of languages with less transparent orthography, such as French or English perform differently in the recalibration paradigm as compared to Dutch speakers. A cross-linguistic comparison, but also an investigation of bi- or multilingual speakers would be interesting in order to disentangle the influence of the nature of a specific language on audiovisual integration.

It has been found that dyslexic readers bypass difficulties in phonological processing by compensating on the level of semantic processing. Therefore, it would be interesting to see whether there is a difference in performance - mainly also compared to non-dyslexic readers - when stimuli consist of real words instead of simple non-words. Of course, this adds additional difficulties for the analysis, since processes of lexical and semantic retrieval need to be taken into consideration. Nevertheless, the use of stimuli that are closer to linguistic information in the everyday life of readers, might be valuable - especially in non-typical reading individuals who necessarily rely even more on contextual information.

Conclusion

It was the goal of the present fMRI study to investigate the neural correlates of text-based and lipread recalibration in dyslexic adults. There was a significant behavioural recalibration effect in both, the text-based and the lipread conditions. However, we did not find a significant difference in the strength of the recalibration effect between the two conditions. This shows that dyslexic participants in our study did not perform differently in the recalibration task depending on whether the disambiguating information was presented in form of written text or in form of video recordings of lip movements. Whole brain group comparisons further revealed a similar neural activation pattern for the text-based and the lipread condition in brain areas typically associated with language processing such as the (posterior) occipital cortex, the superior-temporal cortex, and the pre-motor cortex. There were however differences in brain activation between conditions, which were mostly due to the nature of the visual stimulus (text vs. video). While the lipread condition led to broad activation in the occipital cortex extending to the fusiform face area, the text-based condition elicited stronger activation in the posterior part of the occipital cortex, as well as in the visual word form area. Additionally, we found significant activation in the left IFG for the text-based condition. This area is associated with speech processing in general, however has also been found to be more active in dyslexics employing compensatory strategies during reading. We can therefore conclude that text-based and lipread recalibration effects were comparable in our study, both on a behavioural as well as on a neural level. For the text-based, but not for the lipread condition, we found activation in brain areas typically associated with compensatory mechanisms in dyslexia, indicating the presence of successfully developed compensation strategies, which in turn also explains the comparable behavioural results for the two conditions. However, more research is needed for example by comparing the performance and neural activation patterns of dyslexic to non-dyslexic adults. Only by doing so, we are able to put these findings in place and make meaningful statements about the mechanisms and neural correlates of audiovisual integration in dyslexic readers.

References

Baart, M., de Boer-Schellekens, L., & Vroomen, J. (2012). Lipread-induced phonetic recalibration in dyslexia. *Acta Psychologica*, *140*(1), 91–95. https://doi.org/10.1016/j.actpsy.2012.03.003

Bertelson, P., Vroomen, J., & De Gelder, B. (2003). Visual recalibration of auditory speech identification: A McGurk aftereffect. *Psychological Science*, *14*(6), 592–597. https://doi.org/10.1046/j.0956-7976.2003.psci_1470.x

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. https://doi.org/10.1093/brain/awp308

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. https://doi.org/10.1016/j.cub.2009.01.065

Blomert, L. (2011). The neural signature of orthographic–phonological binding in successful and failing reading development. *Neuroimage*, *57*(3), 695–703. https://doi.org/10.1016/j.neuroimage.2010.11.003

Boersma, P. (2001). Praat, a system for doing phonetics by computer. *Glot. Int.*, *5*(9), 341–345. https://doi.org/https://hdl.handle.net/11245/1.200596

Bonte, M., Correia, J. M., Keetels, M., Vroomen, J., & Formisano, E. (2017). 10.1038/s41598-017-05356-3. *Scientific Reports*, *7*(1), 1–11. https://doi.org/10.1038/s41598-017-05356-3

Brasted, P. J., & Wise, S. P. (2004). Comparison of learning-related neuronal activity in the dorsal premotor cortex and striatum. *European Journal of Neuroscience*, *19*(3), 721–740. https://doi.org/10.1111/j.0953-816X.2003.03181.x

Breman, H., Mulders, J., Fritz, L., Peters, J., Pyles, J., Eck, J., Bastiani, M., Roebroeck, A., Ashburner, J., & Goebel, R. (2020). An image registration-based method for EPI distortion correction based on opposite phase encoding (COPE). *International Workshop on Biomedical Image Registration*, 122–130. https://doi.org/10.1007/978-3-030-50120-4_12

Brookes, S. T., Whitely, E., Egger, M., Smith, G. D., Mulheran, P. A., & Peters, T. J. (2004). Subgroup analyses in randomized trials: risks of subgroup-specific analyses;: power and sample size for the interaction test. *Journal of clinical epidemiology*, *57*(3), 229–236. https://doi.org/10.1016/j.jclinepi.2003.08.009

Brus, B., & Voeten, M. (1991). Een-minuut-test vorm A en B, schoolvorderingstest voor de technische leesvaardigheid bestemd voor groep 4 tot en met 8 van het basisonderwijs. Verantwoording en handleiding. *Lisse: Swets & Zeitlinger.*

Cohen, L., Lehéricy, S., Chochon, F., Lemer, C., Rivaud, S., & Dehaene, S. (2002). Language-specific tuning of visual cortex? Functional properties of the Visual Word Form Area. *Brain*, *125*(5), 1054–1069. https://doi.org/10.1093/brain/awf094

Correia, J. M., Jansma, B. M., & Bonte, M. (2015). Decoding articulatory features from fMRI responses in dorsal speech regions. *Journal of Neuroscience*, *35*(45), 15015–15025. https://doi.org/10.1523/JNEUROSCI.0977-15.2015

Di Folco, C., Guez, A., Peyre, H., & Ramus, F. (2020). Epidemiology of developmental dyslexia: A comparison of DSM-5 and ICD-11 criteria. *medRxiv*. https://doi.org/10.1101/2020.12.18.20248189

Goebel, R., Esposito, F., & Formisano, E. (2006). Analysis of functional image analysis contest (FIAC) data with brainvoyager QX: From single-subject to cortically aligned group general linear model analysis and self-organizing group independent component analysis. *Human brain mapping*, *27*(5), 392–401. https://doi.org/10.1002/hbm.20249

Grill-Spector, K., Knouf, N., & Kanwisher, N. (2004). The fusiform face area subserves face perception, not generic within-category identification. *Nature neuroscience*, *7*(5), 555–562. https://doi.org/10.1038/nn1224

Habib, M. (2020). Developmental dyslexia: A new look at clinical features and brain mechanisms. *Handbook of Clinical Neurology* (pp. 47–59). Elsevier. https://doi.org/10.1016/B978-0-444-64148-9.00004-1

Hanakawa, T., Honda, M., Okada, T., Fukuyama, H., & Shibasaki, H. (2003). Differential activity in the premotor cortex subdivisions in humans during mental calculation and verbal rehearsal tasks: a functional magnetic resonance imaging study. *Neuroscience Letters*, *347*(3), 199–201. https://doi.org/10.1016/S0304-3940(03)00692-X

Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nature reviews neuroscience*, *8*(5), 393–402. https://doi.org/10.1038/nrn2113

Joubert, S., Beauregard, M., Walter, N., Bourgouin, P., Beaudoin, G., Leroux, J.-M., Karama, S., & Lecours, A. R. (2004). Neural correlates of lexical and sublexical processes in reading. *Brain and language*, *89*(1), 9–20. https://doi.org/10.1016/S0093-934X(03)00403-6

Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: a module in human extrastriate cortex specialized for face perception. *Journal of neuroscience*, *17*(11), 4302–4311. https://doi.org/10.1523/JNEUROSCI.17-11-04302.1997

Kearns, D. M., Hancock, R., Hoeft, F., Pugh, K. R., & Frost, S. J. (2019). The neurobiology of dyslexia. *Teaching Exceptional Children*, *51*(3), 175–188. https://doi.org/10.1016/S0166-4115(08)62794-6

Keetels, M., Bonte, M., & Vroomen, J. (2018). A selective deficit in phonetic recalibration by text in developmental dyslexia. *Frontiers in psychology*, *9*, 710. https://doi.org/10.3389/fpsyg.2018.00710

Keetels, M., Schakel, L., Bonte, M., & Vroomen, J. (2016). Phonetic recalibration of speech by text. *Attention, Perception, & Psychophysics*, *78*(3), 938–945. https://doi.org/10.3758/s13414-015-1034-y

Kilian-Hütten, N., Formisano, E., & Vroomen, J. (2017). Multisensory integration in speech processing: neural mechanisms of cross-modal aftereffects. *Neural mechanisms of language* (pp. 105–127). Springer. https://doi.org/10.1007/978-1-4939-7325-5_6

Kilian-Hütten, N., Valente, G., Vroomen, J., & Formisano, E. (2011). Auditory cortex encodes the perceptual interpretation of ambiguous sound. *Journal of Neuroscience*, *31*(5), 1715–1720. https://doi.org/10.1523/JNEUROSCI.4572-10.2011

Kronschnabel, J., Brem, S., Maurer, U., & Brandeis, D. (2014). The level of audiovisual print–speech integration deficits in dyslexia. *Neuropsychologia*, *62*, 245–261.

Liberman, A. M. (1992). The relation of speech to reading and writing. *Advances in psychology* (pp. 167–178). Elsevier.

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, *264*(5588), 746–748. https://doi.org/10.1038/264746a0

Meister, I. G., Wilson, S. M., Deblieck, C., Wu, A. D., & Iacoboni, M. (2007). The essential role of premotor cortex in speech perception. *Current Biology*, *17*(19), 1692–1696. https://doi.org/10.1016/j.cub.2007.08.064

Mirman, D., Chen, Q., Zhang, Y., Wang, Z., Faseyitan, O. K., Coslett, H. B., & Schwartz, M. F. (2015). Neural organization of spoken language revealed by lesion–symptom mapping. *Nature communications*, *6*(1), 1–9. https://doi.org/10.1038/ncomms7762

Murakami, T., Restle, J., & Ziemann, U. (2011). Observation-execution matching and action inhibition in human primary motor cortex during viewing of speech-related lip movements or listening to speech. *Neuropsychologia*, *49*(7), 2045–2054. https://doi.org/10.1016/j.neuropsychologia.2011.03.034

Nergård-Nilssen, T., & Hulme, C. (2014). Developmental dyslexia in adults: Behavioural manifestations and cognitive correlates. *Dyslexia*, *20*(3), 191–207. https://doi.org/10.1002/dys.1477

Ozernov-Palchik, O., & Gaab, N. (2016). Tackling the 'dyslexia paradox': reading brain and behavior for early markers of developmental dyslexia. *Wiley Interdisciplinary Reviews: Cognitive Science*, *7*(2), 156–176. https://doi.org/10.1002/wcs.1383

Price, C. J., & Devlin, J. T. (2003). The myth of the visual word form area. *Neuroimage*, *19*(3), 473–481. https://doi.org/10.1016/S1053-8119(03)00084-3

Raij, T., Uutela, K., & Hari, R. (2000). Audiovisual integration of letters in the human brain. *Neuron*, *28*(2), 617–625. https://doi.org/10.1016/S0896-6273(00)00138-0

Rasamimanana, M., Barbaroux, M., Colé, P., & Besson, M. (2020). Semantic compensation and novel word learning in university students with dyslexia. *Neuropsychologia*, *139*, 107358. https://doi.org/10.1016/j.neuropsychologia.2020.107358

Raschle, N. M., Chang, M., & Gaab, N. (2011). Structural brain alterations associated with dyslexia predate reading onset. *Neuroimage*, *57*(3), 742–749. https://doi.org/10.1016/j.neuroimage.2010.09.055

Romanovska, L., Janssen, R., & Bonte, M. (2019). Reading-induced shifts in speech perception in dyslexic and typically reading children. *Frontiers in psychology*, *10*, 221. https://doi.org/10.3389/fpsyg.2019.00221

Romanovska, L., Janssen, R., & Bonte, M. (2021). 10.1016/j.nicl.2021.102588. *NeuroImage: Clinical*, *30*, 102588. https://doi.org/10.1016/j.nicl.2021.102588

Romanski, L. M. (2012). Integration of faces and vocalizations in ventral prefrontal cortex: implications for the evolution of audiovisual speech. *Proceedings of the National Academy of Sciences*, *109*(Supplement 1), 10717–10724. https://doi.org/10.1073/pnas.1204335109

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. https://doi.org/10.1007/s11682-017-9694-y

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. https://doi.org/10.1207/s1532799xssr0803_6

Seydell-Greenwald, A., Wang, X., Newport, E., Bi, Y., & Striem-Amit, E. (2021). Spoken language comprehension activates the primary visual cortex. *bioRxiv*, 2020–12. https://doi.org/10.1101/2020.12.02.408765

Shaywitz, S. E., & Shaywitz, B. A. (2008). Paying attention to reading: the neurobiology of reading and dyslexia. *Development and psychopathology*, *20*(4), 1329–1349. https://doi.org/10.1017/S0954579408000631

Snowling, M. J., Hulme, C., & Nation, K. (2020). Defining and understanding dyslexia: past, present and future. *Oxford Review of Education*, *46*(4), 501–513. https://doi.org/10.1080/03054985.2020.1765756

Ullas, S., Formisano, E., Eisner, F., & Cutler, A. (2020). Audiovisual and lexical cues do not additively enhance perceptual adaptation. *Psychonomic bulletin & review*, *27*(4), 707–715. https://doi.org/10.3758/s13423-020-01728-5

Ullas, S., Hausfeld, L., Cutler, A., Eisner, F., & Formisano, E. (2020). Neural correlates of phonetic adaptation as induced by lexical and audiovisual context. *Journal of*

Cognitive Neuroscience, *32*(11), 2145–2158. https://doi.org/10.1162/jocn_a_01608

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. https://doi.org/10.1093/cercor/bhl007

Van Linden, S., & Vroomen, J. (2007). Recalibration of phonetic categories by lipread speech versus lexical information. *Journal of Experimental Psychology: Human Perception and Performance*, *33*(6), 1483. https://doi.org/10.1037/0096-1523.33.6.1483

Van Rossum, G., & Drake, F. L. (2009). *Python 3 Reference Manual.* CreateSpace.

Vroomen, J., van Linden, S., De Gelder, B., & Bertelson, P. (2007). Visual recalibration and selective adaptation in auditory–visual speech perception: Contrasting build-up courses. *Neuropsychologia*, *45*(3), 572–577. https://doi.org/10.1016/j.neuropsychologia.2006.01.031

Ward, J. (2019). *The student's guide to cognitive neuroscience.* Routledge. https://doi.org/10.4324/9781351035187

Watkins, K. E., Strafella, A. P., & Paus, T. (2003). Seeing and hearing speech excites the motor system involved in speech production. *Neuropsychologia*, *41*(8), 989–994. https://doi.org/10.1016/S0028-3932(02)00316-0

Weiner, K. S., & Grill-Spector, K. (2012). The improbable simplicity of the fusiform face area. *Trends in cognitive sciences*, *16*(5), 251–254. https://doi.org/10.1016/j.tics.2012.03.003

Yamada, Y., Stevens, C., Dow, M., Harn, B. A., Chard, D. J., & Neville, H. J. (2011). Emergence of the neural network for reading in five-year-old beginning readers of different levels of pre-literacy abilities: an fMRI study. *Neuroimage*, *57*(3), 704–713. https://doi.org/10.1016/j.neuroimage.2010.10.057

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. https://doi.org/10.1371/journal.pone.0110337

Appendix A

Abstract

Everyday life is full of ambiguous speech sounds, and contextual information has been found to facilitate their perception and disambiguation. In individuals with developmental dyslexia, phonological processing and letter-speech sound integration are impaired, leading to difficulties in reading. In this fMRI study, dyslexic adults were exposed to a phonetic recalibration paradigm with ambiguous speech sounds. Additionally, participants were presented simultaneously with either written text or video recordings of lip movements, which served as disambiguating information. The strength of perceptual shift (recalibration) was calculated comparing the two conditions hypothesizing that lip movements would serve as better disambiguating stimulus than text in this specific sample. There was however a significant recalibration effect in both conditions and no significant difference in strength of effect between the two. As hypothesized, brain areas associated with reading and audio-visual integration, such as the visual, auditory, and motor cortex were active in both conditions. Further, there were stimulus-dependent differences in neural activation, with the text-based condition eliciting stronger activation in the posterior occipito-temporal cortex and visual word form area (VWFA) while the lipread condition lead to a broader activation cluster in the occipital cortex including fusiform face area (FFA). The text-based condition further elicited activation in the left inferior frontal gyrus (IFG). Given the significant recalibration effect for both, written text and lip movements and the involvement of the left IFG during text processing, it can be concluded that participants successfully employed compensation strategies during the integration of written text and speech.

Keywords: dyslexia, speech perception, audiovisual integration, phonetic recalibration, fMRI

Appendix B

Zusammenfassung

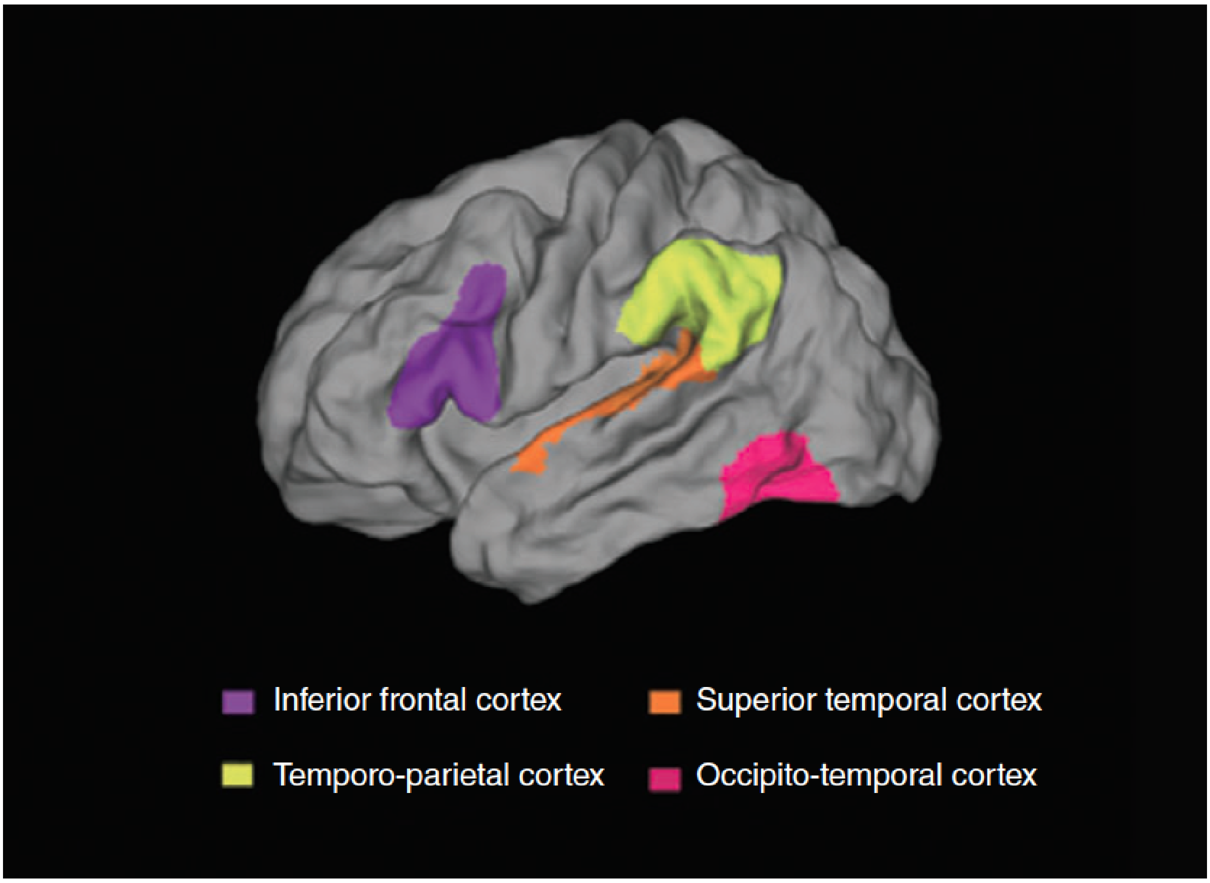
Unser Alltag ist geprägt von ambigen Sprachlauten. Kontextinformationen tragen zu deren erfolgreichen Wahrnehmung und Disambiguierung bei. Personen mit Dyslexie haben Schwierigkeiten, Verknüpfungen zwischen Buchstaben und Sprachlauten herzustellen, sowie phonologische Information zu verarbeiten, was sich negativ auf Leseleistung auswirkt. In dieser fMRT-Studie, nahmen Personen mit Dyslexie and einem phonetischen Rekalibrierungsexperiment teil, in welchem ihnen ambige Sprachlaute präsentiert wurden. Zusätzlich wurden entweder Text oder Videoaufnahmen von Lippenbewegungen als disambiguierende Information gezeigt. Ein Vergleich der Stärke der Wahrnehmungsveränderung des ambigen Lauts (Rekalibrierung) zwischen den beiden Bedingungen wurde erstellt. Dabei wurde erwartet, dass Lippenbewegungen als besserer disambiguierender Stimulus dienen würden als Text. Es wurde jedoch ein signifikanter Rekalibrierungseffekt in beiden Bedingungen und kein signifikanter Unterschied in der Stärke des Effekts zwischen den beiden gefunden. Wie im Vorfeld angenommen, ergab die fMRT-Analyse dass Hirnregionen aktiv waren, die für Lesen und audio-visuelle Integrierung zuständig sind. Dies waren vor allem Areale im visuellen, auditiven, sowie im Motorkortex. Zusätzlich wurden stimulus-abhängige Unterschiede in neuronaler Aktivität zwischen den Bedingungen gefunden, sodass die Text-Bedingung zu stärkerer Aktivierung des posterioren okzipitalen Kortex und der Visual Word Form Area führte während in der Lippenbewegungs-Bedingung stärkere und breiter verteilte Aktivität im okzipitalen Kortex sowie in der Fusiform Face Area gefunden wurden. Zusätzlich war in der Text-Bedingung der linke inferio-frontale Gyrus (IFG) aktiv. Aufgrund des signifikanten Rekalibrierungseffekts in beiden Bedingungen und der Beteiligung des linken IFG in der Text-Bedingung, kann gefolgert werden dass die Versuchspersonen während der Integration von Text und Sprachlauten erfolgreich Kompensationsstrategien angewandt haben.

Appendix C

List of Figures

Figure 1

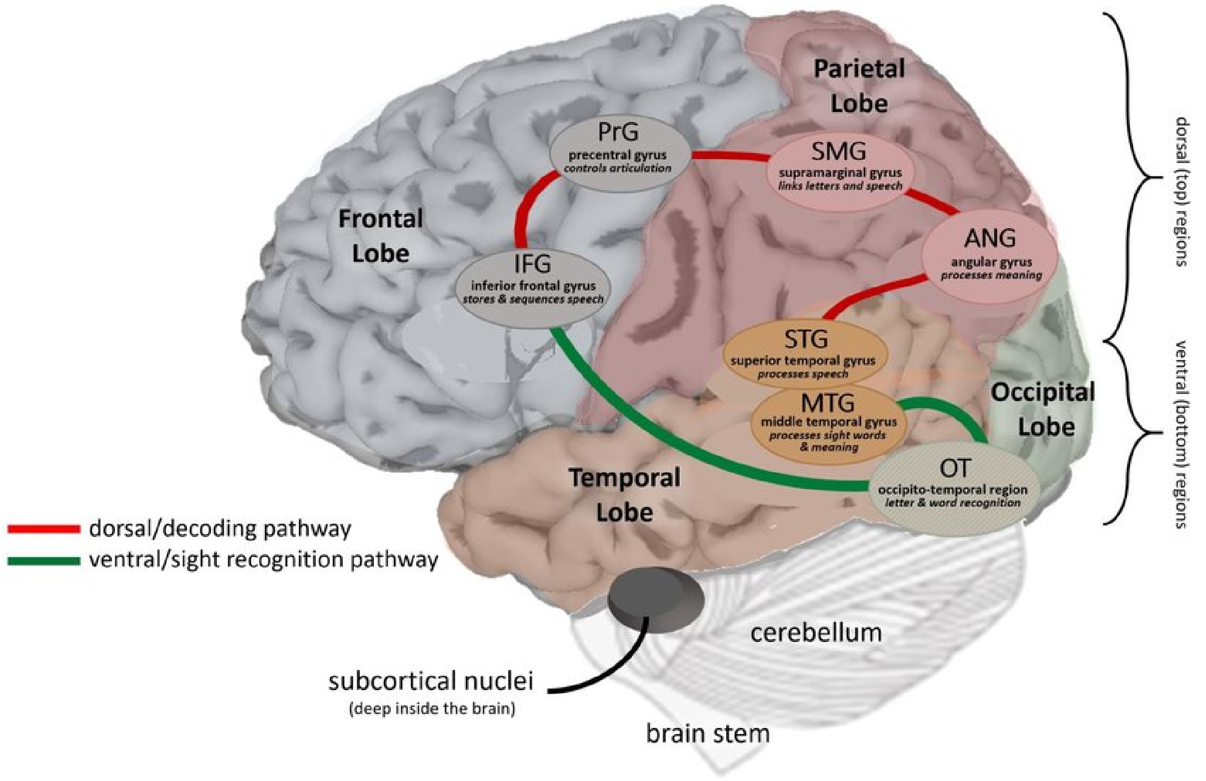
Reading network in the brain impaired in dyslexia



Note. This figure (taken from Ozernov-Palchik & Gaab, 2016) displays brain regions important for reading associated with atypical function or structure in people with dyslexia.

Figure 2

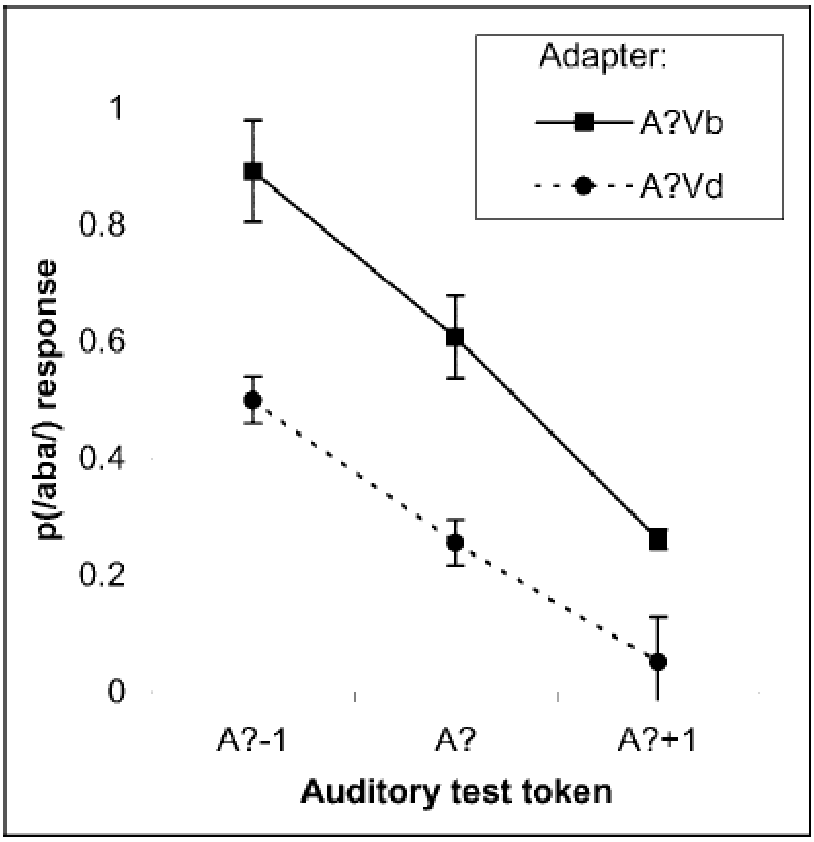
Dorsal and ventral reading pathways in the brain



Note. This figure (taken from Kearns et al., 2019) displays the dorsal decoding (red) and the ventral recognition (green) pathways in the brain that are active during reading.

Figure 3

Results from the post-test in the study by Bertelson et al. (2003)



Note. The figure (taken from Bertelson et al., 2003) displays the proportion of /aba/ judgements (y-axis) to the three post-test sounds (x-axis) after exposure to audiovisual trials with /aba/ (solid line) or /ada/ (dotted line) as disambiguating text.

Figure 4

Experimental design for the scanning session

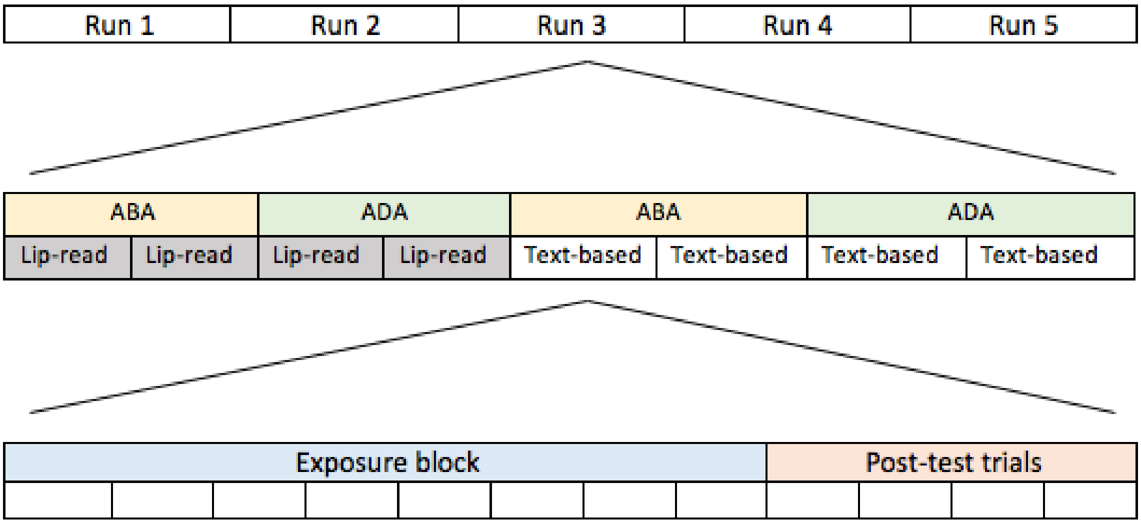
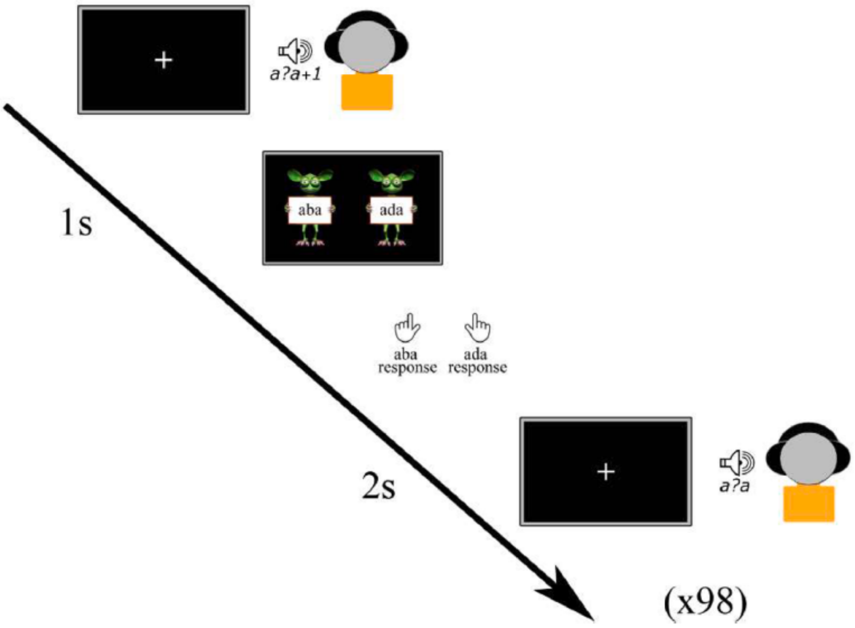


Figure 5

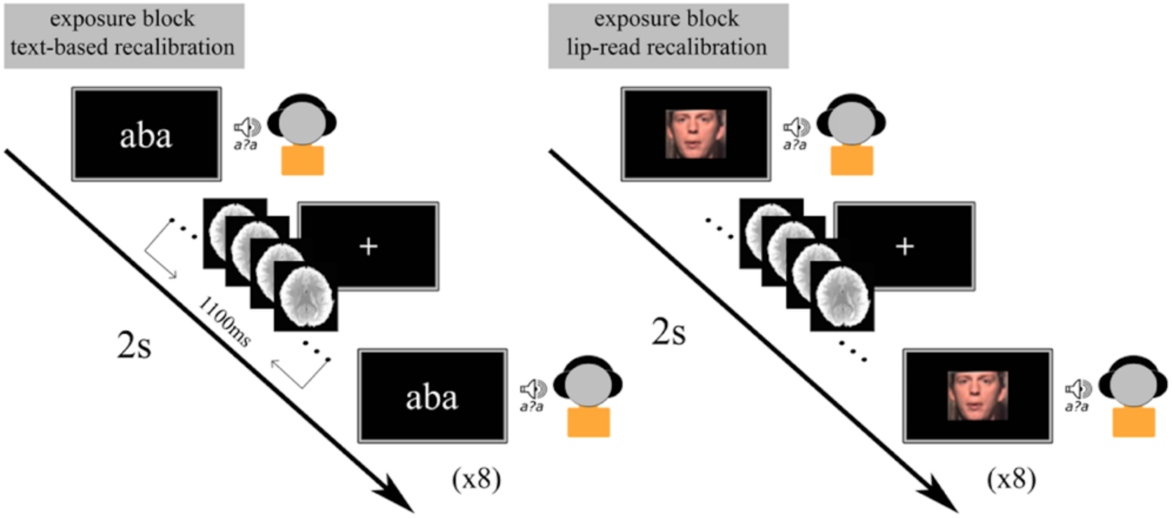
Pre-test



Note. This figure (taken from Romanovska et al., 2021) displays the experimental setup for the pre-test, where the participant is presented with the 9 sound stimuli to determine the individual most ambiguous sound by indicating with a button-press whether the heard sound was "aba" or "ada".

Figure 6

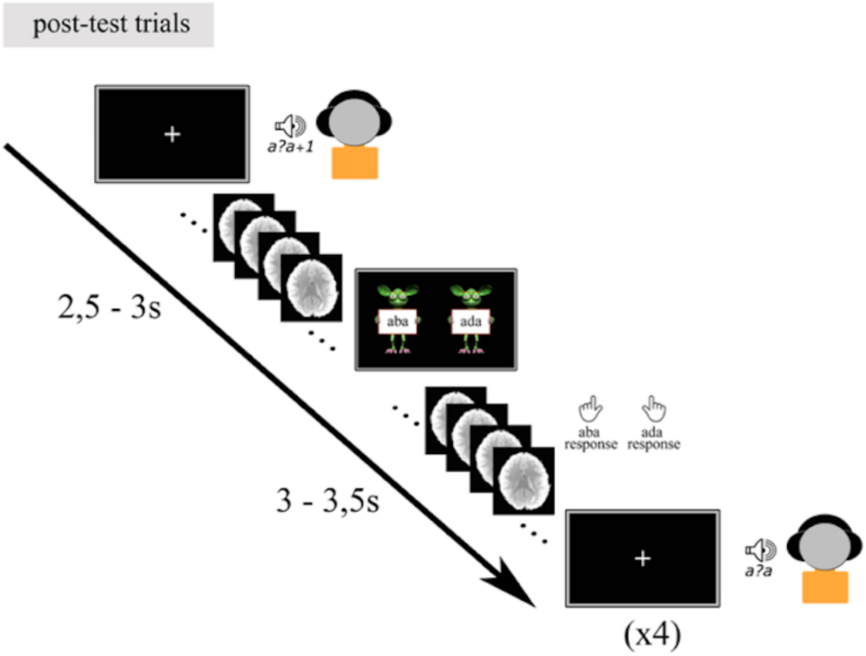
Exposure blocks



Note. This figure displays the experimental setup for the text-based (left) and lipread (right) audiovisual exposure blocks in the MRI scanner.

Figure 7

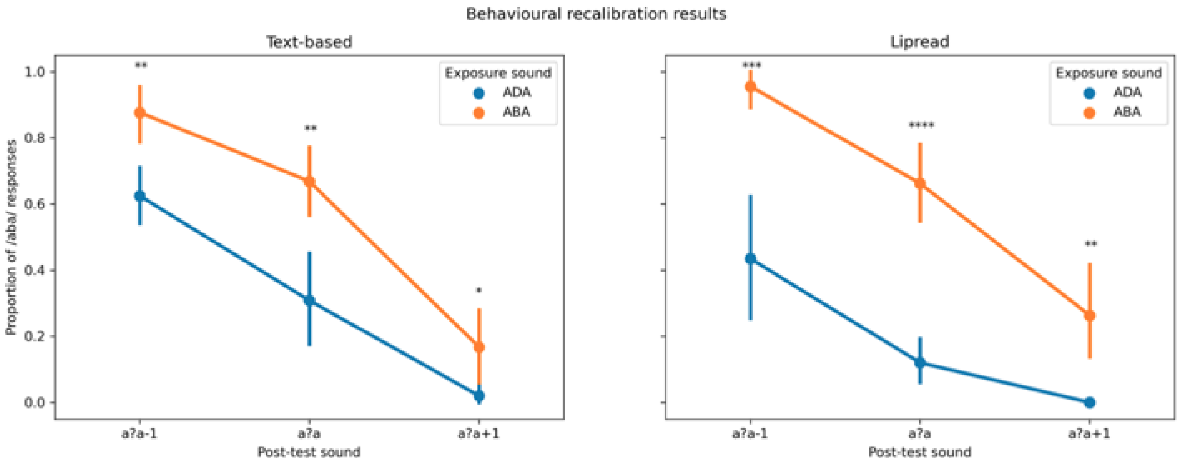
Post-test trials



Note. This figure displays the experimental setup for the post-test trials in the MRI scanner.

Figure 8

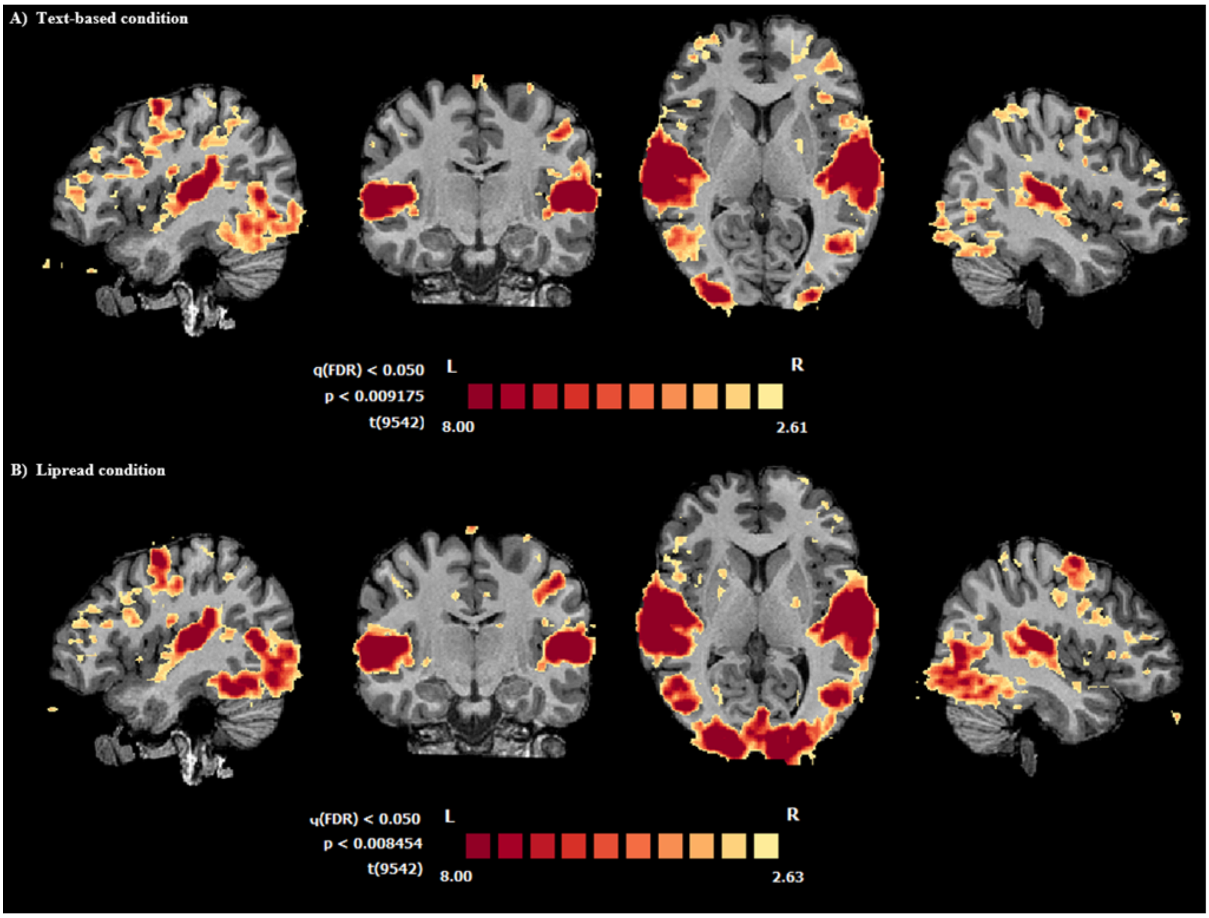
Results of the behavioural recalibration task



Note. This figure displays the results of the behavioural recalibration task in the text-based (left) and the lipread (right) and condition. The y-axis indicates the average proportion for /aba/ responses to the three post-test sounds displayed on the x-axis after "aba" exposure blocks (orange) and "ada" exposure blocks (blue). Vertical bars represent the standard error (* = p<=0.05, ** = p<=0.01, *** = p<=0.001, **** = p<=0.0001).

Figure 9

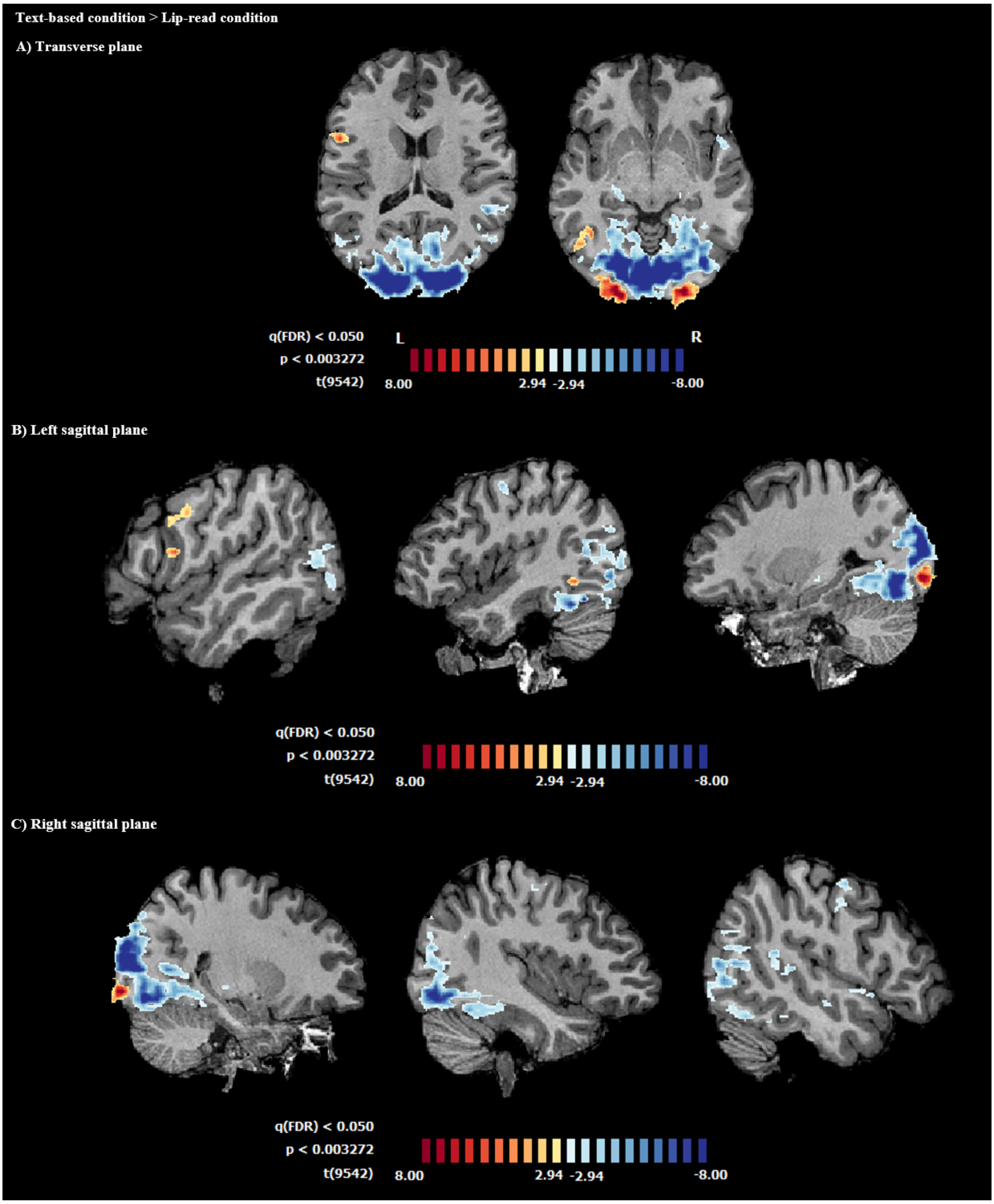
Results of the fMRI analysis for both conditions separately vs. baseline



Note. This figure displays the brain activation pattern during (a) text-based exposure block vs. baseline; (b) lipread exposure block vs. baseline (L = left, R = right).

Figure 10

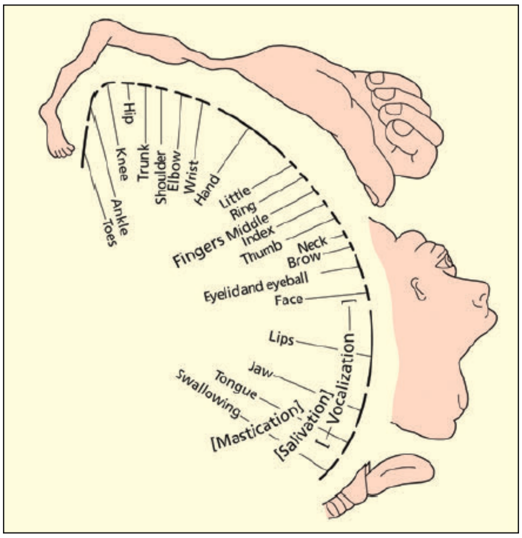
Results of the fMRI analysis for lipread vs. text-based condition



Note. This figure displays the brain activation pattern for lipread exposure block (blue) vs. text-based exposure block (red) in (a) transverse plane; (b) left sagittal plane; (c) right sagittal plane (L = left, R = right).

Figure 11

Somatotopic division of the primary motor cortex



Note. This figure (taken from Ward, 2019) displays the somatotopic division of the primary motor cortex with the respective body parts controlled by a certain area.

Appendix D

List of Tables

Table 1

Age and results from the behavioural reading tests

	Mean	SD	Range
Age	26.3	5.25	18-35
Eén-Minuut-Test (fluency)	75.8	11.3	58-93
Eén-Minuut-Test (accuracy)	74.9	11.1	57-91
Spoonerism task	121.83	44.94	65-212.7

Note. This table shows the mean, standard deviation (SD), and range for participants' age and scores in the Eén-Minuut-Test and the spoonerism task

Table 2

Results from the 2x2x3 repeated-measures ANOVA

	F Value	Num DF	Den DF	Pr > F
Exposure (aba vs. ada)	48.4535	1.00	9.00	0.0001
Post-test sounds	56.2729	2.00	18.00	0.0000
Condition	1.3969	1.00	9.00	0.2675
Exposure (aba vs. ada) : Post-test sounds	7.8822	2.00	18.00	0.0035
Exposure (aba vs. ada) : Condition	8.2195	1.00	9.00	0.0186
Post-test sounds : Condition	3.0807	2.00	18.00	0.0707
Exposure (aba vs. ada) : Post-test sounds : Condition	0.3549	2.00	18.00	0.7061