Tilburg University

Download date: 12. May. 2021

PAPER

WILEY Developmental Science

Audio-visual speech in noise perception in dyslexia

Thijs van Laarhoven | Mirjam Keetels | Lemmy Schakel* | Jean Vroomen

Department of Cognitive Neuropsychology, Tilburg University, Tilburg, The Netherlands

Correspondence
Thijs van Laarhoven, Department of Cognitive Neuropsychology, Tilburg University, Tilburg, The Netherlands.
Email: T.J.T.M.vanLaarhoven @TilburgUniversity.edu

*Present address: Department of Health, Medical and Neuropsychology, Leiden University, Leiden, The Netherlands

Abstract

Individuals with developmental dyslexia (DD) may experience, besides reading problems, other speech-related processing deficits. Here, we examined the influence of visual articulatory information (lip-read speech) at various levels of background noise on auditory word recognition in children and adults with DD. We found that children with a documented history of DD have deficits in their ability to gain benefit from lip-read information that disambiguates noise-masked speech. We show with another group of adult individuals with DD that these deficits persist into adulthood. These deficits could not be attributed to impairments in unisensory auditory word recognition. Rather, the results indicate a specific deficit in audio-visual speech processing and suggest that impaired multisensory integration might be an important aspect of DD.

RESEARCH HIGHLIGHTS

- We report that children with a documented history of developmental dyslexia have deficits in the ability to gain benefit from lip-read information that disambiguates noise-masked speech.
- We show with another group of adult individuals with developmental dyslexia that these deficits persist into adulthood.
- These deficits are unlikely to be attributed to impairments in unisensory auditory word recognition.
- The current results indicate a specific deficit in audio-visual word recognition in our sample of individuals with dyslexia and suggest that impaired multisensory integration might be an important aspect of developmental dyslexia.

1 | INTRODUCTION

The ability to comprehend spoken language is vital for human communication. Although speech is primarily perceived as an auditory experience, input from the visual modality may have a profound influence on speech perception. A textbook example of this is what is commonly referred to as the 'McGurk' effect. In a typical McGurk experiment, the presentation of an auditory recording of the sound /aba/ in conjunction with a visual articulation of the sound /aga/ results in the illusory auditory experience of /ada/ (McGurk & MacDonald, 1976).

The McGurk effect elegantly demonstrates that even incongruent speech sounds and visual lip movements can be combined into unified percepts. A more common example of audio-visual speech integration is the enhancement of speech perception by congruent lip-read information. Observing a speaker's articulatory movements can substantially enhance the ability to comprehend speech, especially in suboptimal listening conditions (MacLeod & Summerfield, 1987; Sumby & Pollack, 1954). The ability to integrate audio-visual speech signals starts to develop during infancy (Hillairet de Boisferon, Tift, Minar, & Lewkowicz, 2016; Lewkowicz, Minar, Tift, & Brandon, 2015; Patterson & Werker, 2003), and continues to improve into late childhood (McGurk & MacDonald, 1976; Ross et al., 2011; Tremblay et al., 2007). Age-related effects on McGurk-type tasks have been widely replicated. Research from Tremblay et al., for instance, has shown that 5- to 9-year-old children are less susceptible to McGurk illusions than older children in the ages 10 to 19. More recently, a study on the development of audio-visual speech perception indicates that the beneficial effect of lip-reading on speech comprehension is roughly 25% lower in children compared to adults (Ross et al., 2011).

While the integration of speech information from multiple modalities is considered as an automatic process that is frequently utilized in everyday life, some individuals struggle to adequately acquire this fundamental ability. Of particular interest are individuals with developmental dyslexia (DD) who, despite normal intelligence, experience difficulties in learning to read. With an estimated worldwide prevalence ranging from 5 to 10%, DD is the most common subtype of

neurodevelopmental disorders (Lyon, Shaywitz, & Shaywitz, 2003; Peterson & Pennington, 2015). Apart from deficits in spelling and writing, individuals with DD may also experience difficulties in the comprehension of speech (American Psychiatric Association, 2000; Hahn, Foxe, & Molholm, 2014; Lyon et al., 2003; World Health Organization, 1992). While this deficit is less noticeable in optimal listening conditions (e.g., speech-in-quiet), it becomes more evident in challenging listening conditions. Several studies have found that speech-in-noise perception is impaired in both children and adults with DD (Boets et al., 2011; Calcus, Colin, Deltenre, & Kolinsky, 2015; Dole, Hoen, & Meunier, 2012; Dole, Meunier, & Hoen, 2014; Ziegler, Pech-Georgel, George, & Lorenzi, 2009).

Findings from studies examining the influence of lip-reading on auditory speech perception in DD are, however, somewhat inconsistent. De Gelder and Vroomen (1998) compared the performance of 9–14-year-old DD children on an audio-visual speech perception task with age- and reading-level matched neurotypical control groups. The task consisted of identifying synthetic speech varying in place of articulation on an acoustic 9-point continuum between /ba/ and /da/. These syllables were either presented in isolation (auditory-only or visual-only) or the sounds were accompanied by the visual articulation of /ba/ or /da/ (audio-visual condition). Participants made a forced-choice decision between 'ba' or 'da'. The influence of visual articulations was measured as the difference in average identification performance between the auditory-only and audio-visual condition. The results showed that individuals with DD were worse than controls in lip-reading, and there was a trend indicating that they were less influenced by visual speech in the audio-visual condition (de Gelder & Vroomen, 1998). Hayes, Tiippana, Nicol, Sams, & Kraus (2003) examined the susceptibility of DD children to the McGurk illusion in three separate conditions: clear (no noise), low-level white Gaussian noise (signal-to-noise ratio [SNR] 0 dB) and high-level white Gaussian noise (SNR -12 dB). DD and neurotypical children identified congruent audio-visual stimuli similarly in all conditions. However, DD children were more likely to report hearing only the visual component of incongruent audio-visual stimuli in high-level background noise than their neurotypical counterparts, thus suggesting that children with DD may actually rely *more* on lip-read information than neurotypical controls.

Another study examined the ability of adults with DD to benefit from lip-read information (Ramirez & Mann, 2005). Natural consonant-vowel stimuli (e.g., /ba/, /da/, /ga/, /ka/, /ma/) were presented in speech-shaped background noise at several SNRs (i.e., 7 dB, −2 dB, and −7 dB) either in isolation or accompanied by visual articulatory cues. The results showed an increased masking effect of background noise for DD subjects in the unisensory condition compared to neurotypical participants. More importantly, DD individuals benefitted *less* from visual articulatory information than their neurotypical counterparts (Ramirez & Mann, 2005). Yet another study examined lip-read-induced phonetic recalibration in adults with DD and found that lip-reading in individuals with DD was in fact as in neurotypical controls (Baart, de Boer-Schellekens, & Vroomen, 2012). In this study, participants were exposed to an ambiguous synthetic speech sound

(falling between /aba/ and /ada/) in combination with clear lip-read speech (i.e., a video of a speaker pronouncing either /aba/ or /ada/). The typical result is that the unimodal ambiguous sound is more likely perceived as /aba/ after exposure to audio-visual /aba/ than after exposure to audio-visual /ada/. This learning effect occurs because, presumably, the conflict between sound and vision is reduced by a shift in the auditory phoneme boundary. The results of Baart et al. (2012) showed that dyslexic adults were as susceptible to the influence of the visual speech adapter as neurotypical controls.

From these studies, it thus appears that there is no consensus about whether individuals with DD have impairments in audio-visual speech processing: several studies reported smaller effects of visual articulatory information, one found larger effects and another found a null-effect. The studies that reported smaller effects used either speech-shaped background noise to mask the speech (Ramirez & Mann, 2005), or synthetic – instead of natural – speech sounds (de Gelder & Vroomen, 1998). This suggests that potential difficulties in audio-visual speech perception experienced by individuals with DD may become more apparent as listening conditions become more challenging. However, additional research is necessary to fully characterize the extent of audio-visual speech processing deficiencies in DD. Of particular importance is to further examine whether the presumed deficit to integrate audio-visual speech-in-noise in DD has a developmental trajectory. Do children and adults with DD have persistent impairments in the integration of audio-visual speech throughout their life, or do these deficits ameliorate with age, as has been reported for other neurodevelopmental disorders, such as autism spectrum disorder (Foxe et al., 2015)?

Apart from this question, it remains to be elucidated whether the apparent deficits in DD in audio-visual speech processing depend on the SNR. In neurotypical adults, speech perception is maximally enhanced by lip-reading at a specific SNR. This SNR is typically located between the extreme SNRs where observers either have to rely completely on lip-reading, or where visual articulatory information is largely redundant due to the high fidelity of the speech sound (Ma, Zhou, Ross, Foxe, & Parra, 2009; Ross, Saint-Amour, Leavitt, Javitt, & Foxe, 2007). Interestingly, this specific SNR seems absent in children. Ross et al. (2011) reported that, unlike 10–11-year-olds and adults, 5–7-year-old children show an overall smoother audio-visual gain curve without a characteristic peak. The absence of an 'optimal' SNR for audio-visual speech integration in young children suggests that the increased sensitivity for lip-read cues at a particular SNR develops somewhere during adolescence. Since DD is a developmental disorder associated with impaired speech-in-noise perception and possible deficiencies in multisensory speech integration abilities, the development of such a window might be altered as well. Therefore, the aim of the present study was to examine the audio-visual speech perception abilities of both children and adults with DD at various SNR values.

We used a spoken word recognition paradigm including various levels of background noise. For the sake of comparability with previous work on audio-visual word recognition in noise in neurotypical children and adults (Ross et al., 2011; Ross et al., 2007), a similar approach was utilized. Dutch mono- and disyllabic high-frequency nouns were

embedded in different levels of pink noise, and were presented either with or without lip-read speech. Speech-in-noise perception abilities of two cohorts of children and adults with DD were compared with two cohorts of individually age-matched neurotypical controls. It was expected that dyslexics would have more difficulty in perceiving uni-sensory speech in noise, and that neurotypical adults would benefit more from the addition of articulatory information compared to children due to a more developed ability to integrate multisensory speech signals (Ross et al., 2011).

2 | METHODS

2.1 | Participants

Four participant groups were included in this study: children with DD, adults with DD, neurotypical children and neurotypical adults. Criteria for typical development (TD) were age-appropriate academic performance and no history of neurological or psychiatric disorders.

Fifteen DD children (7 female, mean age 10.20 years, SD = 1.42) and 15 individually age-matched (i.e., pairwise) TD children (7 female, mean age 10.20 years, SD = 1.42) participated in this study. Children with DD were recruited at a local center for dyslexia and learning disorders (Cognitio, St Willebrord & Roosendaal, the Netherlands). All DD children were currently receiving treatment for reading and spelling problems. Their neurotypical age-matched counterparts were recruited at a local elementary school (St Caecilia, Berkel-Enschot, the Netherlands). Seventeen adults with a documented history of DD (8 female, mean age 20.59 years, SD = 1.91) were recruited from Tilburg University. Some had received specific treatment for DD during childhood or adolescence and all reported persistent problems with reading and spelling. Seventeen individually age-matched TD adults (12 female, mean age 20.59 years, SD = 1.91) took part in the experiment and were recruited from the same university. All participants were native speakers of Dutch, reported normal hearing, and had normal or corrected-to-normal vision. Adults and parents of the children gave written informed consent prior to participating in the study. Children were reimbursed with a small gift for their participation. Adults either received credits in hours as part of a curricular requirement or were reimbursed with €5 for their participation. This study was conducted in accordance with the Declaration of Helsinki (1964). All procedures applied in this experiment were approved by the local ethics committee (EC-2014.45).

2.2 | Stimuli and materials

Stimulus materials have been described before in van der Zande, Jesse, and Cutler (2014). Visual stimuli consisted of video recordings of 120 simple mono- and disyllabic Dutch nouns articulated by a female native Dutch speaker (e.g., 'boom' = tree, 'kamer' = room, 'suiker' = sugar), and a still image of the same speaker exhibiting a neutral facial expression. The recorded videos were digitally remastered so that the length of the video (4 s) and the onset of the speech sound (1.5 s after video onset) were identical across all trials. Viewing

distance was approximately 50 cm. The entire face of the speaker was visible on a neutral background and measured approximately 11.45° horizontally (ear to ear) and 18.33° vertically (hairline to chin). The speech sounds included in the stimulus set contained all ten viseme categories distinguishable in the Dutch language (see van der Zande et al., 2014, for an overview). Visemes are sets of speech sounds produced with similar external articulatory configurations that cannot be distinguished from visual information alone (van Son, Huiskamp, Bosman, & Smoorenburg, 1994). Speech sounds were recorded at a sampling rate of 44.1 kHz and presented over Sennheiser HD201 headphones at a fixed level of approximately 50 dB sound pressure level (SPL) at ear-level.

2.3 | Design

Word recognition performance during each trial was scored dichotomously as correct or incorrect. Only verbal responses that exactly matched the presented nouns were considered correct. Modality and SNR were included as within-subjects factors and AgeGroup (adults, children) and DD (yes, no) as between-subjects factor. Modality was manipulated in two separate conditions: audio-visual (AV) and auditory-only (A). A visual-only condition was not included since previous work using the same stimuli reported very low identification scores in unimodal lip-read word recognition (van der Zande et al., 2014). In the AV condition, speech sounds were presented in conjunction with a video of the speaker articulating the noun. In the A condition, speech sounds were presented while a still image of the speaker's face was displayed. Speech sounds in both conditions were embedded in four levels of pink noise presented at 50, 54, 58 and 62 dB SPL, resulting in SNRs of 0, −4, −8 and −12 dB SPL.

Eight of the 120 nouns included in the stimulus set were selected for practice trials. The remaining 112 nouns were divided into eight subsets of equal size and difficulty. Subset difficulty was based on average viseme overlap (van der Zande et al., 2014) and proportion of disyllabic versus monosyllabic nouns (nine mono syllabic, five disyllabic). Each Modality (AV, A) × SNR (0, −4, −8, −12) combination was assigned to one of the eight subsets (e.g., subset 1: AV, SNR 0, subset 2: AV SNR −4, etc.). Finally, all speech sounds were merged in random order into a single stimulus list and divided into eight equal sized blocks (14 speech sounds in each block). To reduce possible item-specific effects, eight different stimulus lists were generated such that each Modality × SNR combination was assigned once to every subset (e.g., subset 1 List1: AV, SNR 0, subset 1 List2: A, SNR 0, etc.). In addition, stimulus presentation order was randomized across all lists to control for recency and primacy effects.

2.4 | Procedure

Participants were seated in front of the screen and were instructed to attentively listen to the speech sounds and watch the speaker's face. The experimenter ensured that the participant's direction of view was maintained during the experiment and reminded participants when necessary. A practice block of eight trials was included at the start of

the experiment to familiarize participants with the experimental conditions. Total duration of each trial was approximately 8.5 s (Figure 1). Each trial began with the display of a 2050–2150 ms black screen accompanied by a 2000 ms fade-in of pink noise. After a 400 ms fade-in of the first frame the video started. Audio onset occurred 1500 ms after video onset. Each trial concluded 4000 ms after video onset with a 400 ms fade-out of the last video frame accompanied by a 2000 ms fade-out of pink noise and followed by a 1600 ms black screen. After each trial, participants vocally repeated the word they heard. Only responses that exactly matched the presented nouns were considered correct. In cases of an inaudible response, the experimenter initiated the display of a message on the screen requesting the participant to repeat their answer. Responses that remained inaudible after being repeated once were considered as incorrect. Pacing of the trials was under the control of the experimenter. After receiving a clearly audible response, the experimenter initiated the next trial with a button press. Participants were allowed to take a short break after each block to minimize possible fatigue effects. Total duration of the experiment was approximately 20 minutes.

A list of all the nouns included in the experiment was presented after the final block. Participants were instructed to encircle the nouns they did not know the meaning of. Eight pseudowords were intermixed with the regular nouns in the list to control for the possible tendency of participants to report all words as known or not report unknown words as a social desirable response. This test was performed to exclude participants with insufficient vocabulary knowledge.

Two standardized reading tests of Dutch words and pseudowords were used to determine the participant's reading level. The first test (Brus & Voeten, 1973) consists of a list of 116 regular words ordered from lower to higher degree of reading difficulty (e.g., '*been*' = leg, '*verslagenheid*' = consternation). The second test (van den Bos, Lutje Spelberg, Scheepstra, & de Vries, 1994) consists of a list of 116 pseudowords of increasing difficulty (e.g., '*deek*', '*notsberapong*'). Both lists were printed out in four columns and instructions were similar for both tests: participants were asked to read aloud as many words on the list as possible while emphasizing correct pronunciation. The score for each test was calculated as the sum of correctly pronounced words within the predetermined timespan (1 min for the regular word list, 2 min for the pseudoword list).

3 | RESULTS

3.1 | Word recognition performance

Individual proportions of correctly recognized words were calculated for each Modality (AV, A) × SNR (0, −4, −8, −12). Grand average percentages of correct responses as a function of Modality and SNR are shown for DD and TD children and adults separately in Figure 2. Individual audio-visual gain was calculated for each subject as the difference in performance between the unisensory and multisensory condition (i.e., AV−A) across all SNRs, as shown in Figure 3.

A generalized linear mixed-effects model with a logistic linking function to account for the dichotomous dependent variable was fitted to the data (lme4 package in R version 3.2.2). The model included fixed effects for Modality (AV, V), SNR (0, −4, −8, −12), DD (yes, no) and AgeGroup (children, adults). We used the maximal random effect structure supported by the data, with uncorrelated random intercepts and slopes by subjects for all within-subject variables (condition, SNR and their interaction). All categorical factors were recoded such that their values were centered around 0. Hence, the fitted coefficients could be interpreted as the difference in correct responses (in log-odds) between two factor levels (i.e., AV vs. A, DD vs. TD, adults vs. children). The fitted model was: Correct ~ 1 + Modality × SNR × Age Group × DD + (1 + Modality × SNR || Subject). Fixed effect coefficient estimates are shown in Table 1. The model revealed a significant main effect for the intercept ($b = -1.01$, $SE = 0.04$, $p < .001$), indicating an overall bias towards an incorrect response, which fits the overall response distribution (see Figure 2). There were main effects of SNR ($b = 0.99$, $SE = 0.03$, $p < .001$), Modality ($b = -1.60$, $SE = 0.08$, $p < .001$) and AgeGroup ($b = 0.45$, $SE = 0.08$, $p < .001$). The Modality × SNR interaction ($b = 0.39$, $SE = 0.06$, $p < .001$) was significant, indicating that the effect of lip-reading was not uniform across SNRs. The Modality × AgeGroup interaction was also significant ($b = -.49$, $SE = 0.16$, $p < .01$), showing that the effect of lip-reading was different for children and adults. More importantly, a main effect of DD ($b = -0.20$, $SE = 0.08$, $p = .01$) and a Modality × DD interaction were found ($b = 0.37$, $SE = 0.16$, $p = .02$), indicating that the effect of lip-reading was different for DD and neurotypical subjects. There were no other main or interaction effects (all p-values > .05). Several Bonferroni corrected

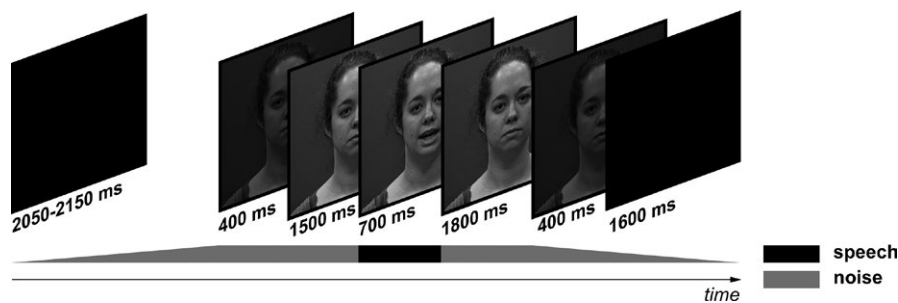**FIGURE 1** Time-course of an audio-visual trial containing the noun '*kamer*' = room. Speech sounds were either presented in conjunction with a video of the speaker articulating the noun (AV) or with a still image of the speaker's face (A). Speech sounds in both conditions were embedded in four levels of pink noise presented at 50, 54, 58 and 62 dB SPL, resulting in signal-to-noise ratios of 0, −4, −8 and −12 dB SPL

TABLE 1 Fixed effect coefficients and standard errors for the fitted mixed effects regression model: Correct ~ 1 + Modality × SNR × AgeGroup × DD + (1 + Modality × SNR || Subject)

| Fixed factor	Estimate	Standard error	z-value	p
(Intercept)	−1.01	0.04	−24.61	<.001***
Modality	−1.60	0.08	−20.47	<.001***
DD	−0.20	0.08	−2.49	.01*
AgeGroup	0.45	0.08	5.46	<.001***
SNR	0.99	0.03	30.99	<.001***
DD × AgeGroup	0.10	0.16	0.47	.54
DD × SNR	−0.03	0.06	−0.44	.66
DD × Modality	0.37	0.16	2.32	.02*
Modality × AgeGroup	−0.49	0.16	−3.11	<.01**
Modality × SNR	0.39	0.06	6.11	<.001***
SNR × AgeGroup	0.04	0.06	0.59	.55
DD × Modality × AgeGroup	0.40	0.32	1.26	.21
DD × Modality × SNR	0.02	0.13	0.19	.85
DD × SNR × AgeGroup	0.01	0.13	0.10	.92
AgeGroup × Modality × SNR	0.19	0.13	1.47	.14
DD × Modality × AgeGroup × SNR	−0.25	0.25	−1.00	.32

*p < .05; **p < .01; ***p < .001.
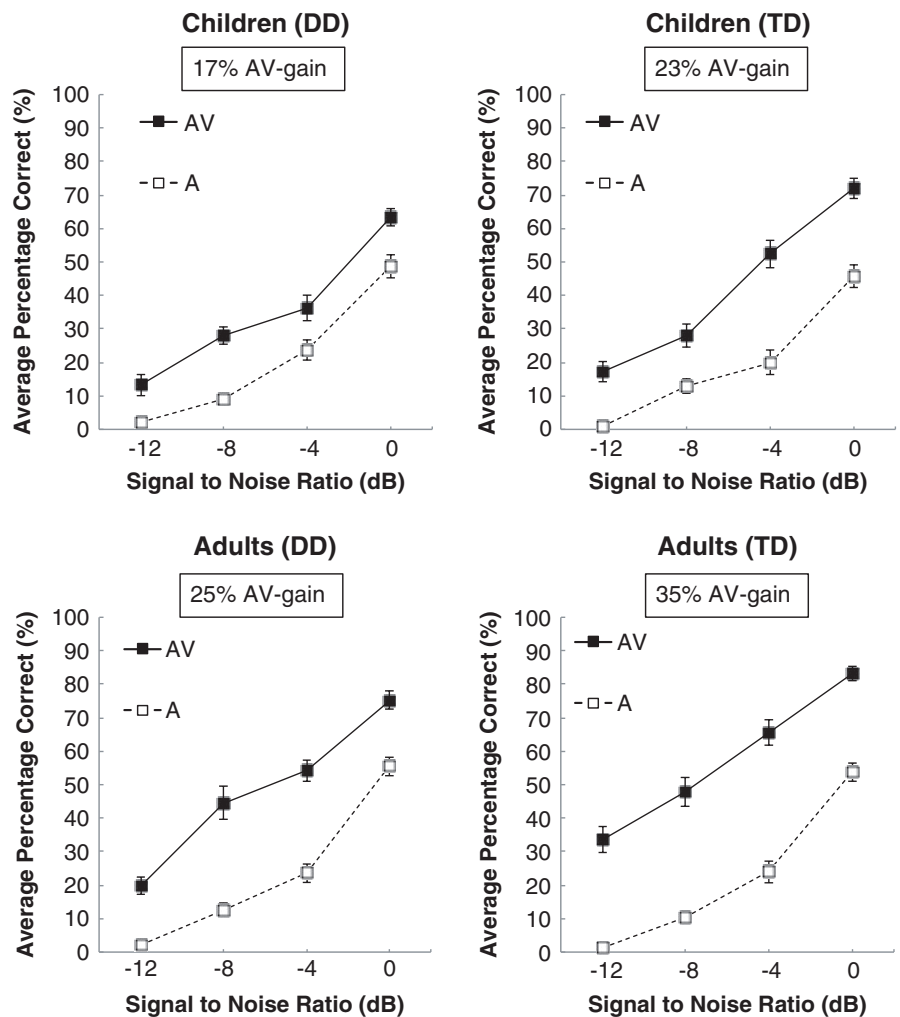
FIGURE 2 Grand average word recognition performance as a function of signal-to-noise ratio (SNR) and Modality. Average word recognition performance (% correct) at each SNR is plotted separately for children and adults with developmental dyslexia (DD) and typical development (TD). Error bars represent one standard error of the mean

paired samples *t*-tests (two-tailed) were performed to further explore the observed interaction effects.

3.1.1 | Modality × DD

Further analysis of the Modality × DD interaction revealed that DD subjects benefited substantially less from lip-reading than neurotypical participants (−8%, $b = −.40$, $SE = 0.09$, $p < .0001$), but did not perform differently from TD participants in the unisensory condition ($p > .90$). The scatter plot in Figure 3 provides a more detailed depiction of the difference in AV gain between TD and DD subjects in both AgeGroups.

3.1.2 | Modality × AgeGroup

Across all SNRs, both children (+20%, $b = 1.36$, $SE = 0.12$, $p < .0001$) and adults (+30%, $b = 1.85$, $SE = 0.11$, $p < .0001$) were able to distinguish more words in the lip-reading condition compared to the unisensory condition. As expected, adults benefited more from lip-reading than children (+10%, $b = −.69$, $SE = 0.09$, $p < .0001$). There were no significant differences in unisensory performance between children and adults ($p > .12$). Figure 3 illustrates the gradual linear improvement in audio-visual gain with age in both TD and DD subjects.

3.1.3 | Modality × SNR

Across all subjects, performance at each SNR was higher when lip-read information was available ($p < .0001$ for all contrasts), while a decrease in SNR resulted in lower recognition performance in

Individual AV Gain by Age

FIGURE 3 Scatter plot of the individual audio-visual (AV) gain as a function of age in years. AV gain was calculated for each participant as the difference in performance between the audio-visual and auditory-only condition (AV−A) across all signal-to-noise ratios. Explained variance (R^2) of the linear regression lines for participants with developmental dyslexia (DD) and typical development (TD) are displayed

each Modality ($p < .0001$ for all contrasts). We further explored the Modality × SNR interaction by examining the absolute performance difference across all subjects between the unisensory and multisensory condition (i.e., AV−A) at each SNR. Gain by lip-reading increased from 24% to 30% between the SNRs 0 and −4 dB and decreased from 27% to 20% at the lowest SNRs of −8 and −12 dB. Unisensory performance at the peak of audio-visual gain (i.e., SNR −4 dB) was approximately 23%.

3.2 | Vocabulary

The mean percentage of unknown words was calculated for each participant and submitted to a two-way analysis of variance (ANOVA) with DD and AgeGroup as between-subjects factors. The average percentage of unknown words across all subjects was low ($M = 3.74$, $SD = 4.82$), indicating that participants' vocabulary size was sufficient for the experiment. Analysis revealed a significant main effect of AgeGroup ($F(1, 60) = 40.36$, $p < .001$, $\eta_p^2 = .40$), indicating that the percentage of words that subjects were unfamiliar with was 6% lower in adults ($M = 0.89$, $SD = 1.31$) than in children ($M = 6.96$, $SD = 5.31$). There was no main effect of DD and no AgeGroup × DD interaction effect, indicating that the amount of unknown words did not differ between DD and TD participants.

3.3 | Reading fluency

Regular and pseudoword reading fluency scores were analyzed with a multivariate ANOVA with DD and AgeGroup as between-subjects factors. A significant interaction between DD and AgeGroup was found for reading fluency of regular words ($F(1, 60) = 5.59$, $p = .02$, $\eta_p^2 = .09$). This interaction was further explored with simple main effects tests on each level of DD and AgeGroup. The simple main effect test for DD on each AgeGroup level revealed a main effect of DD for children ($F(1, 28) = 25.30$, $p < .001$, $\eta_p^2 = .48$) and adults ($F(1, 32) = 6.43$, $p = .02$, $\eta_p^2 = .17$). DD children ($M = 44.73$, $SD = 16.66$) read significantly fewer words during the regular word-reading test compared to age-matched TD children ($M = 74.73$, $SD = 15.99$). DD adults ($M = 85.29$, $SD = 15.83$) also performed significantly lower on the regular word-reading test compared to age-matched TD adults ($M = 97.41$, $SD = 11.72$). The simple main effect test for AgeGroup on each DD level showed that both DD children ($F(1, 30) = 49.80$, $p < .001$, $\eta_p^2 = .62$) and TD children ($F(1, 30) = 21.28$, $p < .001$, $\eta_p^2 = .4$) had lower regular word-reading test scores than their adult counterparts. There were main effects of AgeGroup ($F(1, 60) = 102.38$, $p < .001$, $\eta_p^2 = .63$) and DD ($F(1, 60) = 57.85$, $p < .001$, $\eta_p^2 = .49$) on pseudoword reading fluency. Pseudoword reading fluency was lower in DD participants ($M = 54.47$, $SD = 26.52$) compared to neurotypical subjects ($M = 82.66$, $SD = 22.06$). Children ($M = 48.40$, $SD = 22.94$) read significantly fewer pseudowords in the predetermined timespan than adults ($M = 86.35$, $SD = 18.58$). The results on the reading tests thus confirmed that reading fluency developed with age and that both cohorts of DD individuals had persistent reading problems.
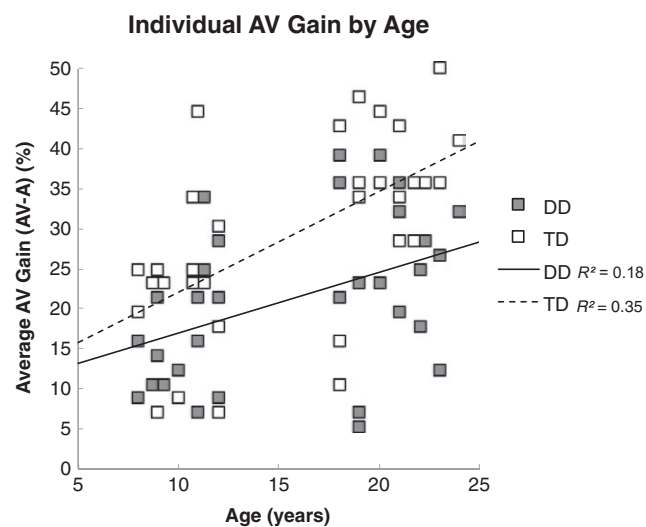
4 | DISCUSSION

This study examined the audio-visual speech perception abilities in DD. Two cohorts of children and adults with DD and two TD age-matched control cohorts were included in a single experiment. A spoken word recognition paradigm was used in which a large set of different simple mono- and disyllabic nouns embedded in variable levels of pink noise was presented either with or without visual articulatory information.

4.1 | Differences between DD and TD subjects

This study showed that individuals with DD benefited substantially less than TD subjects from lip-read information that disambiguates noise-masked speech, regardless of age and SNR. The current results are in line with previous findings published by de Gelder and Vroomen (1998) and Ramirez and Mann (2005), who reported that the processing of synthetic and natural audio-visual consonant-vowel stimuli is atypical in children and adults with DD. The current data extend these findings by showing that the processing of natural audio-visual speech may be atypical in DD individuals as well. Another crucial aspect of the present study is the inclusion of two age cohorts of DD and TD individuals. This enabled us to investigate whether or not the impact of the deficits in audio-visual speech processing is ameliorated during adulthood. We found that the relative impairment of the ability to gain benefit from lip-reading in DD is present in children and adults, despite the fact that the current cohort of DD adults consisted of students with a highly educated background. Taken together, these findings indicate that in DD, despite adequate education and unlike in other neurodevelopmental disorders such as autism spectrum disorder (Brandwein et al., 2013; Foxe et al., 2015; Stevenson, Segers, Ferber, Barense, & Wallace, 2014) and developmental language disorders (Meronen, Tiippana, Westerholm, & Ahonen, 2013), deficits in the processing of audio-visual speech-in-noise do not resolve during adolescence, but persist into adulthood. The current findings are therefore consistent with previous longitudinal studies indicating that dyslexia is a persistent disorder and not merely a condition of transient 'developmental lag' (Francis, Shaywitz, Stuebing, Shaywitz, & Fletcher, 1996; Scarborough, 1984).

Somewhat surprisingly, DD subjects did not show deficits in unisensory auditory word recognition. This is in contrast with previous studies that showed that DD is associated with impaired phonological and basic auditory processing (Hämäläinen, Salminen, & Leppänen, 2013; Navas, Ferraz Ede, & Borges, 2014; Steinbrink, Klatte, & Lachmann, 2014). One might argue that, since the adults included in this study have all reached university-level education, they could have had compensatory cognitive resources (e.g., semantic knowledge, verbal ability and visual memory) that minimized the impact of their phonological deficits (Shaywitz et al., 2003). However, the absence of unisensory performance differences between children and adults indicates that all participants, regardless of DD, age and education, were capable of adequately performing the unisensory task. Hence, the cognitive demands for this task probably did not require the aid of compensatory mechanisms. Another potential explanation for the lack

of unisensory auditory speech perception deficits in the current sample of DD individuals might be methodological differences between the current study and previous work. In previous studies reporting speech-in-noise perception deficits in DD, either speech sounds similar to the stimuli included in the present study were used that were embedded in a different type of background noise (i.e., words embedded in babble speech, fluctuating speech-shaped noise or stationary noise) (Boets et al., 2011; Dole et al., 2012; Dole et al., 2014), or both the type of speech sound and background noise were different from those applied in the present study (i.e., consonant-vowel stimuli embedded in babble speech, stationary speech-shaped noise or modulated speech-shaped noise) (Calcus et al., 2015; Ziegler et al., 2009). Importantly, speech-in-noise deficits were not observed in DD subjects when stationary noise was used as a speech masker in a word recognition task similar to the present study (Boets et al., 2011; Dole et al., 2012). The current results are in line with these findings, and suggest that individuals with DD are more sensitive to informational masking induced by speech-shaped noise than energetic masking induced by stationary noise. In addition, individuals with DD seem to experience fewer difficulties in discriminating whole words in noise compared to noise-masked syllables, suggesting that their phonological deficits reside at the sub-word level. However, further research is needed to investigate the impact of both the nature of the task and the type of background noise on the difficulties encountered by DD individuals in the perception of speech-in-noise.

Nevertheless, the fact that DD subjects benefited substantially less from lip-reading is unlikely to be attributed to impairments in unisensory auditory word recognition. The current results therefore indicate a specific deficit in audio-visual word recognition in our sample of individuals with a documented history of DD. Similar results have been found in research on letter-speech sound integration in DD (Blau et al., 2010; Blau, van Atteveldt, Ekkebus, Goebel, & Blomert, 2009; Froyen, Willems, & Blomert, 2011; Mittag, Thesleff, Laasonen, & Kujala, 2013). A functional magnetic resonance imaging (fMRI) study by Blau et al. (2009) investigated the neural processing of letters and speech sounds in DD adults. Their data showed a reduced activation of the superior temporal cortex in DD adults during the integration of letters and speech sounds, which is indicative of impaired audio-visual integration. Another fMRI study (Blau et al., 2010) revealed that DD children differ from TD children in the neural integration of basic letter–speech sound pairs. In TD children, cortical responses were modulated by letter–speech sound congruency, as indicated by the strong suppression effects in the processing of incongruent letter–speech sound pairs. In DD children, however, no such modulation was observed, indicating deficient integration of letters and speech sounds (Blau et al., 2010). An electrophysiological study by Froyen et al. (2011) showed that the development of automatic associations between orthographic tokens and phonemic utterances is impaired in children with DD, despite intact auditory processing abilities. Mittag et al. (2013) found similar results in DD adults, indicating that this failure to acquire automatic audio-visual word associations is persistent into adulthood. These and the present results suggest that a general dysfunction in multisensory integration might be an important aspect of DD.

If a general deficit in multisensory integration is in fact the underlying cause of DD, one could expect that the impact of this impairment might be reduced by explicit interventions, preferably during early childhood. Interestingly, there is some evidence that audio-visual training might indeed improve reading skills in DD (Ecalle, Magnan, Bouchafa, & Gombert, 2009; Fraga Gonzalez et al., 2015; Kast, Baschera, Gross, Jancke, & Meyer, 2011; Kujala et al., 2001). These training programs are mainly focused on increasing phonemic awareness and explicit learning of letter–sound associations, and typically involve tasks such as the discrimination of voicing pairs (e.g., 'ba' versus 'pa') and matching of spoken to written syllables (Hahn et al., 2014). However, all but one (Kujala et al., 2001) of these studies used training programs based on linguistic stimuli. Thus, although the results of studies on audio-visual training in DD certainly look promising, additional research including semantically unrelated stimuli is needed to determine whether the impairment in audio-visual integration, as observed in the current study and previous work, stems from a purely linguistic or more general deficit in multisensory processing.

4.2 | Differences between children and adults

The current findings replicate previous work from Ross et al. (2011), and show that adults experienced more multisensory enhancement by lip-reading than children. Although vocabulary size was slightly lower in children compared to adults, there were no significant differences in unisensory recognition performance between both age groups. Therefore, it is unlikely that the observed difference in audio-visual gain between children and adults can be accounted for by developmental differences in lexicon. Given the evidence from previous research (Foxe et al., 2015; Ross et al., 2011), it seems likely that the ability to benefit from lip-reading continues to develop into late childhood both as a function of exposure to audio-visual speech and as a function of the self-production of speech. In addition, there is ample evidence that substantial developmental changes in cognitive functioning occur throughout childhood due to maturational changes in the brain (Fair et al., 2008; Liston, Matalon, Hare, Davidson, & Casey, 2006; Shaw et al., 2008; Somerville & Casey, 2010). These developmental changes are, to some degree, reflected in the lower test performance scores of children compared to adults on both the regular and pseudoword-reading test. The current findings therefore provide further evidence for the assumption that a combination of developmental, behavioral and environmental influences leads to an increased ability to integrate audio-visual speech signals into unified percepts.

4.3 | Effect of SNR

The masking effect of decreasing SNR values seemed to be more pronounced in the current experiment compared to previous studies (Ma et al., 2009; Ross et al., 2011; Ross et al., 2007). In the present study, the SNR where subjects hardly recognized any words in the A-only condition was located at −12 dB. In previous work, however, unisensory performance accuracy at this SNR ranged from 11 to 20% (Ma et al., 2009; Ross et al., 2011; Ross et al., 2007). In addition, the current data show a peak in audio-visual gain at an SNR of −4 dB whereas in previous work, audio-visual gain by lip-reading peaked at a considerably lower SNR of approximately −12 dB (Ma et al., 2009; Ross et al., 2011; Ross et al., 2007). A possible explanation for this difference in effect of SNR is that the listening configuration used in the present experiment is different from the one applied in previous research. In the studies from Ross et al. (2007, 2011) and Ma et al. (2009), speech sounds were played from a centrally located speaker, while the noise was presented from two lateral speakers. Since the sound sources were physically separated from one another, it is possible that subjects used spatial and/or head shadow cues to segregate the speech sounds from the background noise. This well-studied phenomenon, commonly referred to as 'spatial-release from masking' or 'spatial unmasking', is known to significantly improve speech recognition in noisy environments (Bronkhorst, 2000; Darwin, 2008; Grange & Culling, 2016; Hawley, Litovsky, & Culling, 2004). In the present experiment, however, speech sounds and background noise were presented binaurally through headphones and thus the two sound sources appeared as collocated. This prevented the occurrence of spatial-release from masking, which in turn may have resulted in more challenging listening conditions when the SNR was lowered compared to previous research.

4.4 | CONCLUSION

The present study replicated previous findings by Ross et al. (2011), showing that adults experience more multisensory enhancement by lip-reading than children. Most importantly, we found that subjects with a documented history of DD have deficits in their ability to gain benefit from lip-read information in a word recognition in noise task and that these deficits are independent of SNR and persist into adulthood. These deficits could not be attributed to impairments in unisensory word recognition. The current results therefore indicate a specific deficit in audio-visual word recognition in our sample of individuals with a documented history of DD. Additional research is needed to determine whether this deficit stems from a linguistic or more general impairment in multisensory integration.

REFERENCES

American Psychiatric Association (2000). *Diagnostic and statistical manual of mental disorders* (4th ed., text rev.). Washington, DC: Author.

Baart, M., de Boer-Schellekens, L., & Vroomen, J. (2012). Lipread-induced phonetic recalibration in dyslexia. *Acta Psychologica*, 140, 91–95.

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, 868–879.

Blau, V., van Atteveldt, N., Ekkebus, M., Goebel, R., & Blomert, L. (2009). Reduced neural integration of letters and speech sounds links phonological and reading deficits in adult dyslexia. *Current Biology*, 19, 503–508.

Boets, B., Vandermosten, M., Poelmans, H., Luts, H., Wouters, J., & Ghesquiere, P. (2011). Preschool impairments in auditory processing and speech perception uniquely predict future reading problems. *Research in Developmental Disabilities*, 32, 560–570.

Brandwein, A.B., Foxe, J.J., Butler, J.S., Russo, N.N., Altschuler, T.S., Gomes, H., & Molholm, S. (2013). The development of multisensory integration in high-functioning autism: High-density electrical mapping and psychophysical measures reveal impairments in the processing of audiovisual inputs. *Cerebral Cortex*, 23, 1329–1341.

Bronkhorst, A.W. (2000). The cocktail party phenomenon: A review of research on speech intelligibility in multiple-talker conditions. *Acustica*, 86, 117–128.

Brus, B.T.H., & Voeten, M.J.M. (1973). *Een-Minuut-Test. Vorm A en B. Verantwoording en handleiding*. Nijmegen: Berkhout.

Calcus, A., Colin, C., Deltenre, P., & Kolinsky, R. (2015). Informational masking of speech in dyslexic children. *Journal of the Acoustical Society of America*, 137, EL496–502.

Darwin, C.J. (2008). Listening to speech in the presence of other sounds. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 363, 1011–1021.

de Gelder, B., & Vroomen, J. (1998). Impaired speech perception in poor readers: Evidence from hearing and speech reading. *Brain and Language*, 64, 269–281.

Dole, M., Hoen, M., & Meunier, F. (2012). Speech-in-noise perception deficit in adults with dyslexia: Effects of background type and listening configuration. *Neuropsychologia*, 50, 1543–1552.

Dole, M., Meunier, F., & Hoen, M. (2014). Functional correlates of the speech-in-noise perception impairment in dyslexia: An MRI study. *Neuropsychologia*, 60, 103–114.

Ecalle, J., Magnan, A., Bouchafa, H., & Gombert, J.E. (2009). Computer-based training with ortho-phonological units in dyslexic children: New investigations. *Dyslexia*, 15, 218–238.

Fair, D.A., Cohen, A.L., Dosenbach, N.U., Church, J.A., Miezin, F.M., Barch, D.M., ... Schlaggar, B.L. (2008). The maturing architecture of the brain's default network. *Proceedings of the National Academy of Sciences of the United States of America*, 105, 4028–4032.

Foxe, J.J., Molholm, S., Del Bene, V.A., Frey, H.P., Russo, N.N., Blanco, D., ... Ross, L.A. (2015). Severe multisensory speech integration deficits in high-functioning school-aged children with autism spectrum disorder (ASD) and their resolution during early adolescence. *Cerebral Cortex*, 25, 298–312.

Fraga Gonzalez, G., Zaric, G., Tijms, J., Bonte, M., Blomert, L., & van der Molen, M.W. (2015). A randomized controlled trial on the beneficial effects of training letter-speech sound integration on reading fluency in children with dyslexia. *PLoS ONE*, 10, e0143914.

Francis, D.J., Shaywitz, S.E., Stuebing, K.K., Shaywitz, B.A., & Fletcher, J.M. (1996). Developmental lag versus deficit models of reading disability: A longitudinal, individual growth curves analysis. *Journal of Educational Psychology*, 88, 3–17.

Froyen, D., Willems, G., & Blomert, L. (2011). Evidence for a specific cross-modal association deficit in dyslexia: An electrophysiological study of letter-speech sound processing. *Developmental Science*, 14, 635–648.

Grange, J.A., & Culling, J.F. (2016). The benefit of head orientation to speech intelligibility in noise. *Journal of the Acoustical Society of America*, 139, 703–712.

Hahn, N., Foxe, J.J., & Molholm, S. (2014). Impairments of multisensory integration and cross-sensory learning as pathways to dyslexia. *Neuroscience and Biobehavioral Reviews*, 47, 384–392.

Hämäläinen, J.A., Salminen, H.K., & Leppänen, P.H. (2013). Basic auditory processing deficits in dyslexia: Systematic review of the behavioral and event-related potential/ field evidence. *Journal of Learning Disabilities*, 46, 413–427.

Hawley, M.L., Litovsky, R.Y., & Culling, J.F. (2004). The benefit of binaural hearing in a cocktail party: Effect of location and type of interferer. *Journal of the Acoustical Society of America*, 115, 833–843.

Hayes, E.A., Tiippana, K., Nicol, T.G., Sams, M., & Kraus, N. (2003). Integration of heard and seen speech: A factor in learning disabilities in children. *Neuroscience Letters*, 351, 46–50.

Hillairet de Boisferon, A., Tift, A.H., Minar, N.J., & Lewkowicz, D.J. (2016). Selective attention to a talker's mouth in infancy: Role of audiovisual temporal synchrony and linguistic experience. *Developmental Science*, doi:10.1111/desc.12381

Kast, M., Baschera, G.M., Gross, M., Jancke, L., & Meyer, M. (2011). Computer-based learning of spelling skills in children with and without dyslexia. *Annals of Dyslexia*, 61, 177–200.

Kujala, T., Karma, K., Ceponiene, R., Belitz, S., Turkkila, P., Tervaniemi, M., & Näätänen, R. (2001). Plastic neural changes and reading improvement caused by audiovisual training in reading-impaired children. *Proceedings of the National Academy of Sciences of the United States of America*, 98, 10509–10514.

Lewkowicz, D.J., Minar, N.J., Tift, A.H., & Brandon, M. (2015). Perception of the multisensory coherence of fluent audiovisual speech in infancy: Its emergence and the role of experience. *Journal of Experimental Child Psychology*, 130, 147–162.

Liston, C., Matalon, S., Hare, T.A., Davidson, M.C., & Casey, B.J. (2006). Anterior cingulate and posterior parietal cortices are sensitive to dissociable forms of conflict in a task-switching paradigm. *Neuron*, 50, 643–653.

Lyon, G.R., Shaywitz, S.E., & Shaywitz, B.A. (2003). A definition of dyslexia. *Annals of Dyslexia*, 53, 1–14.

Ma, W.J., Zhou, X., Ross, L.A., Foxe, J.J., & Parra, L.C. (2009). Lip-reading aids word recognition most in moderate noise: A Bayesian explanation using high-dimensional feature space. *PLoS ONE*, 4, e4638.

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, 264, 746–748.

MacLeod, A., & Summerfield, Q. (1987). Quantifying the contribution of vision to speech perception in noise. *British Journal of Audiology*, 21, 131–141.

Meronen, A., Tiippana, K., Westerholm, J., & Ahonen, T. (2013). Audiovisual speech perception in children with developmental language disorder in degraded listening conditions. *Journal of Speech, Language, and Hearing Research*, 56, 211–221.

Mittag, M., Thesleff, P., Laasonen, M., & Kujala, T. (2013). The neurophysiological basis of the integration of written and heard syllables in dyslexic adults. *Clinical Neurophysiology*, 124, 315–326.

Navas, A.L., Ferraz Ede, C., & Borges, J.P. (2014). Phonological processing deficits as a universal model for dyslexia: Evidence from different orthographies. *Codas*, 26, 509–519.

Patterson, M.L., & Werker, J.F. (2003). Two-month-old infants match phonetic information in lips and voice. *Developmental Science*, 6, 191–196.

Peterson, R.L., & Pennington, B.F. (2015). Developmental dyslexia. *Annual Review of Clinical Psychology*, 11, 283–307.

Ramirez, J., & Mann, V. (2005). Using auditory-visual speech to probe the basis of noise-impaired consonant-vowel perception in dyslexia and auditory neuropathy. *Journal of the Acoustical Society of America*, 118, 1122–1133.

Ross, L.A., Molholm, S., Blanco, D., Gomez-Ramirez, M., Saint-Amour, D., & Foxe, J.J. (2011). The development of multisensory speech perception continues into the late childhood years. *European Journal of Neuroscience*, 33, 2329–2337.

Ross, L.A., Saint-Amour, D., Leavitt, V.M., Javitt, D.C., & Foxe, J.J. (2007). Do you see what I am saying? Exploring visual enhancement of speech comprehension in noisy environments. *Cerebral Cortex*, *17*, 1147–1153.

Scarborough, H.S. (1984). Continuity between childhood dyslexia and adult reading. *British Journal of Psychology*, *75*, 329–348.

Shaw, P., Kabani, N.J., Lerch, J.P., Eckstrand, K., Lenroot, R., Gogtay, N., … Wise, S.P. (2008). Neurodevelopmental trajectories of the human cerebral cortex. *Journal of Neuroscience*, *28*, 3586–3594.

Shaywitz, S.E., Shaywitz, B.A., Fulbright, R.K., Skudlarski, P., Mencl, W.E., Constable, R.T., … Gore, J.C. (2003). Neural systems for compensation and persistence: Young adult outcome of childhood reading disability. *Biological Psychiatry*, *54*, 25–33.

Somerville, L.H., & Casey, B.J. (2010). Developmental neurobiology of cognitive control and motivational systems. *Current Opinion in Neurobiology*, *20*, 236–241.

Steinbrink, C., Klatte, M., & Lachmann, T. (2014). Phonological, temporal and spectral processing in vowel length discrimination is impaired in German primary school children with developmental dyslexia. *Research in Developmental Disabilities*, *35*, 3034–3045.

Stevenson, R.A., Segers, M., Ferber, S., Barense, M.D., & Wallace, M.T. (2014). The impact of multisensory integration deficits on speech perception in children with autism spectrum disorders. *Frontiers in Psychology*, *5*, 379.

Sumby, W.H., & Pollack, I. (1954). Visual contribution to speech intelligibility in noise. *Journal of the Acoustical Society of America*, *26*, 212–215.

Tremblay, C., Champoux, F., Voss, P., Bacon, B.A., Lepore, F., & Theoret, H. (2007). Speech and non-speech audio-visual illusions: A developmental study. *PLoS ONE*, *2*, e742.

van den Bos, K.P., Lutje Spelberg, H.C., Scheepstra, A.J.M., & de Vries, J.R. (1994). *De Klepel, een test voor de leesvaardigheid van pseudowoorden. Verantwoording, handleiding, diagnostiek en behandeling*. Nijmegen: Berkhout.

van der Zande, P., Jesse, A., & Cutler, A. (2014). Hearing words helps seeing words: A cross-modal word repetition effect. *Speech Communication*, *59*, 31–43.

van Son, N., Huiskamp, T.M.I., Bosman, A.J., & Smoorenburg, G.F. (1994). Viseme classifications of Dutch consonants and vowels. *Journal of the Acoustical Society of America*, *96*, 1341–1355.

World Health Organization (1992). *The ICD-10 classification of mental and behavioural disorders: Clinical descriptions and diagnostic guidelines*. Geneva: Author.

Ziegler, J.C., Pech-Georgel, C., George, F., & Lorenzi, C. (2009). Speech-perception-in-noise deficits in dyslexia. *Developmental Science*. *12*, 732–745.