Perceptual Adaptation to Noise-Vocoded Speech by Lip-Read Information: No Difference between Dyslexic and Typical Readers

Faezeh Pourhashemi, Martijn Baart* and Jean Vroomen
Department of Cognitive Neuropsychology, Tilburg University, 5000 LE Tilburg, the Netherlands
*Corresponding author; e-mail: m.baart@tilburguniversity.edu
ORCID iDs: Pourhashemi: 0000-0003-1885-5763; Baart: 0000-0002-5015-4265; Vroomen: 0000-0001-5923-5988

Received 3 May 2023; accepted 3 May 2024; published online 23 May 2024; published in print 29 May 2024

Abstract
Auditory speech can be difficult to understand but seeing the articulatory movements of a speaker can drastically improve spoken-word recognition and, on the longer-term, it helps listeners to adapt to acoustically distorted speech. Given that individuals with developmental dyslexia (DD) have sometimes been reported to rely less on lip-read speech than typical readers, we examined lip-read-driven adaptation to distorted speech in a group of adults with DD (N = 29) and a comparison group of typical readers (N = 29). Participants were presented with acoustically distorted Dutch words (six-channel noise-vocoded speech, NVS) in audiovisual training blocks (where the speaker could be seen) interspersed with audio-only test blocks. Results showed that words were more accurately recognized if the speaker could be seen (a lip-read advantage), and that performance steadily improved across subsequent auditory-only test blocks (adaptation). There were no group differences, suggesting that perceptual adaptation to disrupted spoken words is comparable for dyslexic and typical readers. These data open up a research avenue to investigate the degree to which lip-read-driven speech adaptation generalizes across different types of auditory degradation, and across dyslexic readers with decoding versus comprehension difficulties.

Keywords
auditory learning, dyslexia, lip-reading, speech adaptation
1. Introduction

Developmental dyslexia (DD) is one of the most prevalent developmental disorders and affects about 5–12% of the population (Eden et al., 2016). It is characterized by difficulties with learning to read and spell despite adequate intelligence and levels of education. Mounting evidence suggests that individuals with DD have less-pronounced letter–speech sound associations than typical readers, next to commonly observed phonological speech processing difficulties (e.g., Blomert, 2011; Ramus, 2003; Vellutino et al., 2004). Given that reading relies on linking letters to sounds, it is plausible that more general disruptions in integrating auditory and visual information underlie DD. There is indeed evidence that individuals with DD do not always benefit from lip-read speech to the same extent as typical readers do, especially when auditory speech is difficult to understand (de Gelder and Vroomen, 1998; Hahn et al., 2014; Ramirez and Mann, 2005; Rüsseler et al., 2015, 2018; Schaadt et al., 2019; van Laarhoven et al., 2018). However, the finding that dyslexics rely less on lip-read information is not always observed (for a recent review, see Pulliam et al., 2023). Here, we examined a so far untested function of lip-read speech, namely in guiding adaptation to acoustically distorted speech.

Speech can sometimes be difficult to understand due to ambiguities in the signal that arise from, for example, background noise or ideosyncratic speaker properties (such as unfamiliar accents). Yet, the perceptual system flexibly adapts to these ambiguities (e.g., Greenspan et al., 1988) when the listener receives additional information that disambiguates the compromised speech input. This disambiguating context can generate an error signal that, in turn, drives changes in the acoustic–phonetic mapping (for a review, see Guediche et al., 2014). For example, adaptation is enhanced when distorted speech is (1) accompanied by a written version of the speech signal (Francis et al., 2007; Loebach et al., 2010; Schwab et al., 1985), (2) preceded by a clear undistorted acoustic version of the speech input (Davis et al., 2005; Hervais-Adelman et al., 2008), or (3) combined with lip-read speech (Pourhashemi et al., 2022). In all these cases, listeners rely on the disambiguating context to more readily adapt to the acoustic distortions and, as a result, thus become better listeners.

Given that listeners with DD may rely less on lip-read speech than typical readers, we examined whether lip-read information can guide adaptation to distorted speech in DD to the same extent as it does in typical readers. A recent study by Gabay and Holt (2021) examined this question using written text rather than lip-read speech as corrective feedback signal. Participants in this study were presented with severely distorted words during a pretest and a posttest phase. In an intervening training phase, less severely distorted words were accompanied by text that informed listeners which word they had heard. This training led to improved word recognition at posttest, but the gain was
less for listeners with DD than for the comparison group. This finding then raises the question if listeners with DD have less adaptation when instead of written text, *lip-read speech* is used as a signal that disambiguates the acoustic message.

The comparison between lip-read speech and text is of importance because in previous studies on ‘phonetic recalibration’ (a form of perceptual learning as explained below), it was found that lip-read speech and text acted differently in listeners with or without DD (Baart *et al*., 2012; Keetels *et al*., 2018). Phonetic recalibration relies, like speech adaptation, on error minimization but its scope is at the acoustic–phonetic mapping of a single phoneme rather than complete words or sentences. More specifically, in the seminal study on lip-read-driven phonetic recalibration (Bertelson *et al*., 2003), listeners were exposed to a video of a speaker who pronounced either /aba/ or /ada/ while an ambiguous speech sound halfway between /aba/ and /ada/ was heard. In auditory-only posttests, identification of the ambiguous sound was shifted toward the previously seen lip-read information, so that the same test sound was perceived more likely as /aba/ when the previous exposure contained lip-read /aba/, and more likely as /ada/ when the previous exposure contained lip-read /ada/. The rationale behind this assimilative effect is that the discrepancy between the heard and seen information is minimized by shifting the auditory phonetic boundary toward the lip-read stimulus. This then leads to assimilative aftereffects in auditory-only posttests (i.e., the phonetic system is ‘recalibrated’). Other research has demonstrated that not only lip-read speech, but also text (letters; see Keetels *et al*., 2018) and lexical information (e.g., Norris *et al*., 2003) can drive phonetic recalibration (see Ullas *et al*., 2022, for a review). Most relevant for the present study is that in individuals with DD, phonetic recalibration by lip-read speech was comparable for dyslexic and typical readers (Baart *et al*., 2012), whereas recalibration was smaller for dyslexic than typical readers when instead of lip-read speech text was used to drive recalibration (Keetels *et al*., 2018). This thus suggests that text-to-sound conversion might be negatively impacted by DD, whereas lip-read-to-sound conversion is not.

Here, we examined speech adaptation to acoustically distorted words by lip-read information in DD. On the one hand, one might expect similar adaptation effects for readers with DD and typical readers given that their lip-read-induced phonetic recalibration to phonemes appears to be comparable (Baart *et al*., 2012; Keetels *et al*., 2018). Alternatively, individuals with DD might show weaker adaptation effects as they may not benefit from lip-read speech to the same extent as typical readers do (de Gelder and Vroomen, 1998; Ramirez and Mann, 2005; Rüsseler *et al*., 2015, 2018; Schaad et al., 2019; van Laarhoven *et al*., 2018) or because there is more a general problem in the
procedural learning system (Nicolson and Fawcett, 2010) that transcends to weaker adaptation effects (Gabay and Holt, 2021).

To examine this question, we used the stimuli and design of Pourhashemi et al. (2022) where multiple audio-only test blocks were interspersed with audiovisual training blocks. Each block contained 15 unique auditory distorted words that were not previously encountered in the task. During audiovisual training, participants heard the distorted words and simultaneously saw the speaker producing it, whereas during audio-only test blocks they only heard distorted words. In this study with typical readers, it was found that accuracy on audiovisual training items was ~65%, whereas participants who — instead of dynamic lip-read information — received a static image of the speaker during training only identified ~21% of the words correctly. Moreover, accuracy on auditory-only test items (presented without any visual information) increased with ~20% when comparing the first with the last test block for participants in the audiovisual lip-read training condition, which starkly contrasted the ~3% increase in performance for participants who had received a static image of the speaker’s face during the audiovisual training blocks (Pourhashemi et al., 2022). In the current study, we only compared, for dyslexic and typical readers, the extent to which lip-read speech boosts spoken-word recognition (i.e., a lip-read advantage of audiovisual training blocks over auditory-only test blocks), and the extent to which listeners adapt to noise-vocoded speech over time (i.e., an increase in performance in subsequent training and test blocks), so without a condition with a static face.

2. Material and Methods

2.1. Participants

Fifty-eight native Dutch Tilburg University students with normal hearing and (corrected to) normal vision were recruited. Twenty-nine students were registered as being dyslexic [i.e., they had provided an official dyslexia declaration (see Note 1) while enrolled at Tilburg University], and the other 29 students were fluent readers from the Tilburg University psychology program (the comparison group). Participants were either paid (10 Euro/h), or they received course credits in return for their participation. The mean age was 20.27 years (SD = 2.03) in the dyslexic group (23 females, 6 males) and 19.75 years (SD = 1.40) in the comparison group (25 females, 4 males). The experiment was conducted in accordance with the Declaration of Helsinki. Prior to participation, all participants provided written informed consent and the experiment was approved by the Tilburg University Ethics Review Board (project ID: EC-2016.48).
2.2. Stimuli

The stimulus set contained 120 audiovisual recordings of mono- and bisyllabic words (see also Pourhashemi et al., 2022, where the full list of items is provided). A male native Dutch speaker (one of the authors: MB) pronounced the words while being recorded by a Nikon D7200 camera and an external microphone attached to the camera (Røde VideoMicro). The videos showed the entire face of the speaker from the neck upward, and were framed as headshots against a black background. The auditory stream was extracted from the video using Adobe Premiere Pro (13.0) and then manipulated in the Praat software (Boersma, 2001). The ‘Shannon-AM-noise’ script written by Chris Darwin (2005) was used to transform the audio into so-called six-channel noise-vocoded speech (NVS). Each auditory signal was decomposed into six nonoverlapping frequency bands: 50–229 Hz, 229–558 Hz, 558–1161 Hz, 1161–2265 Hz, 2265–4290 Hz, and 4290–8000 Hz. Next, the amplitude envelope for each individual band was extracted and combined with a Hann band-pass filtered white-noise signal. The filter smoothing value was defined as the highest frequency divided by 10. All bands were then recombined to create six-channel NVS, in which the combined individual filtered bands did overlap at the frequency band boundaries. Next, each NVS item was mixed onto the original video of the speaker to create audiovisual (AV) stimuli (to use for training blocks as explained below).

2.3. Materials

Participants lip-read-driven auditory learning of NVS was investigated via a spoken-word identification task, and reading performance was assessed via two reading tasks.

2.3.1. Spoken-Word Identification

Participants were seated in front of a full HD monitor (BenQ XL 2540-B, 24.5 inch, 240 Hz refresh rate) in a sound-attenuated booth and were instructed to listen to speech sounds while attentively looking at the screen. A live camera feed from the experimental booth to the experimenter was monitored to assess looking behavior, and participants were instructed (via an intercom) to keep looking at the screen if necessary. The auditory NVS items were delivered through headphones (Sennheiser HD 203) at an intensity of ∼65 dBA (measured at ear level). The experiment was run in the E-prime 3.0 software (available at https://pstnet.com), and lasted for ∼25 min. The monitor in the testing booth (visible to the participant) displayed full-screen lip-read videos during training blocks and a question mark after each training and test trial signaled the participant to provide a response. A second monitor was placed outside of the booth and displayed information that was relevant to the researcher only. It displayed the trial number, the information that identified
the particular block in the experiment, and an orthographic representation of
the stimulus that was presented to the participant. On any given trial, partici-
pants provided a vocal response (while they were seeing the question mark) by
repeating out loud what they had perceived. This response was recorded with a
microphone in the booth (Samson, C01U Pro, USB studio condenser), and also
relayed in real-time to the experimenter via the intercom system. The exper-
imenter then compared the participants’ verbal response to the orthographic
representation on the screen and scored the item as ‘correct’ or ‘incorrect’.
Whenever a response was unclear, the experimenter asked the participant to
repeat the response until the experimenter was confident about the (in)correct
nature of the participant’s response. The next trial automatically started when
the experimenter’s scoring response (correct/incorrect) was collected.

All 120 NVS words were delivered once during the experiment. The words
were divided into eight blocks containing 15 items each. The first block was an
auditory-only familiarization block to acquaint the participants with the task
and stimuli. This was followed by another auditory-only block that represented
base-line performance prior to audiovisual (AV) training (i.e., T1, for Test
block 1). Next, three AV training blocks were delivered (Training 1, Training
2, and Training 3) that were interspersed with auditory-only test blocks labeled
as T2, T3, and T4, respectively (see Fig. 1). Although we refer to the AV trial
blocks as ‘training blocks’ — as in Pourhashemi et al. (2022) — please note
that participants did not receive any corrective feedback. During AV training
blocks, participants heard an NVS word while seeing a video of the speaker
producing that word. The words that were assigned to each block were fixed,
but item order within each block was randomized across participants. Block
order was counterbalanced across participants, such that for example a block
was presented as the auditory test block for some participants while for the
other participants, the same block was presented as an audiovisual training
block. The auditory familiarization block contained the same items for all par-
ticipants, but item order was randomized.

![Figure 1](https://example.com/figure1.png)

**Figure 1.** Experimental design. The familiarization block and Test blocks T1, T2, T3 and T4 consisted of 15 auditory NVS words. Training blocks 1, 2, and 3 consisted of 15 NVS words that were presented in combination with a dynamic face (i.e., lip-read information). After each training or test item, participants repeated aloud what they had heard.
2.3.2. Reading Task

Two standardized reading tests were administered to evaluate participants’ reading skills: The “Een Minuut Test” (“One Minute Test”, Version B; Brus and Voeten, 1979) and the “De Klepel” (Version B; van den Bos et al., 1999). The “Een-Minuut-test” contains an ordered list of 116 words that increase in difficulty, and participants are required to read as many items as they can (out loud) in the provided order, while the experimenter scores any mistakes on a scoring sheet. The “De Klepel” test is similar in structure and procedure, but it contains pseudowords rather than words. Although the official time-limit for that test is two minutes (it is also used to assess reading skill in children), we administered a one-minute version. For both tests, the test score reflects the total number of items that were read within the time-limit, minus the number of items that contained a reading error. Participants’ reading skills were assessed after the spoken-word identification task, always starting with the “Een-Minuut-test” (real words).

2.4. Analysis

To analyze the experimental data, we used the same approach as in Pourhashemi et al. (2022). Specifically, we fitted generalized linear mixed-effects models of the binomial family on the individual data by maximum likelihood estimation (Laplace approximation) using the logit link function. The data were analyzed in R Studio (version 4.10) using the lme4 package (version 1.1-27, Bates et al., 2015). One model was used to analyze the data from the auditory test blocks, and another model was used to analyze the data from the A V training blocks. For the test blocks, the following model was fitted on the individual responses: ‘Correct ~ 1 + Test block × Group + (1|Subject) + (1|Items)’, which included fixed effects for ‘Test block’, and ‘Group’, as well as their interaction effects, with correlated maximal random effects of ‘Subject’ and ‘Item’. In this model, ‘Correct’ (correct or incorrect response) was considered as the dependent variable. The fixed factors ‘Test block’ and ‘Group’ were coded into numeric variables symmetrically around 0 (using contrast coding): ‘Test block’ was recoded into four values (−1.5, −0.5, +0.5, and +1.5, for T1 to T4, respectively) and ‘Group’ was recoded as −1 for the dyslexic group, and +1 for the comparison group. The fitted coefficient for ‘Group’ is interpreted as the difference in correct responses (in log-odds) between the dyslexic and comparison groups. For further analysis, we used the functions in the lsmeans package (Version 2.30-3) in R.

The data from the A V training blocks were analyzed in a similar model, in which — instead of ‘Test block’ — the factor ‘Training block’ was included (three levels, coded as −1 for Training 1, 0 for Training 2, and 1 for Training 3).
Figure 2. Proportions of correct responses across test and training blocks. The data points connected by lines represent the mean proportion of correctly recognized words for auditory-only Test blocks (T1, T2, T3, T4) and audiovisual Training blocks (Training 1, Training 2, Training 3) for the dyslexic readers and the comparison group. Error bars represent one standard error of the mean. The dots represent the individual data.

3. Results

For each training and test block, we computed the proportion of correct responses across items and participants, separately for the dyslexic and comparison group. As can be seen in Fig. 2, adding lip-read information during audiovisual training blocks substantially improved spoken-word recognition relative to audio-only test blocks (= lip-read advantage), performance steadily inclined during subsequent test and training blocks (= adaptation), and all averages were highly comparable across both groups, which was confirmed in our analyses.

3.1. Auditory-Only Test Blocks

A main effect of intercept was observed ($b = -0.53$, SE = 0.26, $p = 0.04$), indicating that the overall response distribution was biased toward incorrect rather than correct responses (see Section 2.4 Analysis for how to interpret other variables; see also Pourhashemi et al., 2022). More importantly, a significant main effect of Test block was found ($b = 0.37$, SE = 0.04, $p < 0.001$), indicating that the proportion of correct responses steadily increased from the first to the last block (i.e., by 16%). The mean increase in performance between the first and last test block (i.e., T4 − T1) was 19% for the DD group, and 13% for the comparison group, with no significant main effect of Group ($b = 0.04$, SE = 0.08, $p = 0.62$), and no significant interaction between Group and Test block ($b = 0.07$, SE = 0.04, $p = 0.08$; see Table 1).

To assess the null effect of Group in more detail, we conducted a Bayesian repeated-measures ANOVA in the JASP software (JASP Team, 2020). All
models that contained the factor Group (either as a main effect or as an interaction term) yielded BF01 factors > 4.80, which can be interpreted as substantial evidence in support of the null effect of Group. The full summary table is presented in Supplementary Table S1.

### 3.2. Audiovisual Training Blocks

There was a main effect of intercept ($b = 2.18$, SE = 0.26, $p < 0.001$), indicative of an overall trend toward providing correct responses, rather than incorrect ones (which contrasts the data from the auditory Test blocks). The main effect of Training block was significant, which means performance significantly increased across subsequent audiovisual training blocks ($b = 0.28$, SE = 0.07, $p < 0.001$) for all comparisons. Performance in both groups from Training block 1 to Training block 3 increased with 6% with no effect of Group ($p > 0.74$; see Table 2). No main effect of Group was observed ($b = 0.05$, SE = 0.10, $p = 0.60$). Again, the null effect of Group was assessed in more detail with a Bayesian repeated-measures ANOVA. All models that contained the factor Group (either as a main effect or as an interaction term) yielded BF01 factors > 11.66, providing no basis to reject the null-hypothesis. The full summary table is presented in Supplementary Table S2.
3.3. Gain by Lip-Read Speech

To examine the extent to which lip-read speech boosted auditory word recognition, we computed ‘lip-read gain’ (see Supplementary Table S3) by subtracting performance in auditory-only test blocks (T1–T4) from the audiovisual training blocks (Training 1–Training 3). On average, adding lip-read speech increased accuracy by 33% with no difference between the DD and comparison groups ($b = 0.01$, SE = 0.03, $p = 0.61$). We also examined if the lip-read gain could explain adaptation (i.e., are good lip-readers also good adapters?). We conducted a linear regression model in which we entered adaptation (quantified as the difference between auditory test blocks T4 and T1) as the dependent variable, and lip-read gain and Group as the predictors.

Table 3 provides a summary of these analyses. As can be seen, there was a significant effect of Intercept ($b = 0.12$, SE = 0.04, $p = 0.001$), indicating that, overall, adaptation was larger than 0.

More importantly, neither Group nor lip-read gain correlated significantly with adaptation ($b = 0.013$, SE = 0.10, $p = 0.21$ and $b = 0.01$, SE = 0.04, $p = 0.68$, respectively). No interaction effect of Group and lip-read gain was observed ($b = 0.03$, SE = 0.10, $p = 0.79$).

Finally, we compared the reading scores on the word- and pseudoword-reading tasks across the groups. Average word- and pseudoword-reading scores were 101.68 and 65.11 for the comparison group, versus 81.83 and 46.17 for the DD group, and on both tasks, the comparison group had outperformed the DD group ($t$ values $> 5.15$, $p$-values $< 0.001$).

4. Discussion

We sought to determine whether lip-read-driven adaptation to noise-vocoded speech is modulated by dyslexia. Although reading scores were — unsurprisingly — lower for the DD group than for the comparison group, we found

Table 3.
Regression models: Adaptation ~ Group* Lip-read gain

<table>
<thead>
<tr>
<th>Predictors</th>
<th>Estimate</th>
<th>SE</th>
<th>t-value</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>0.12</td>
<td>0.04</td>
<td>3.28</td>
<td>0.001***</td>
</tr>
<tr>
<td>Group</td>
<td>0.13</td>
<td>0.10</td>
<td>1.24</td>
<td>0.21</td>
</tr>
<tr>
<td>Lip-read gain</td>
<td>0.01</td>
<td>0.04</td>
<td>0.40</td>
<td>0.68</td>
</tr>
<tr>
<td>Group* Lip-read gain</td>
<td>0.03</td>
<td>0.10</td>
<td>0.26</td>
<td>0.79</td>
</tr>
</tbody>
</table>

* $p < 0.05$; **, $p < 0.01$; ***, $p < 0.001$. RSA: Response Surface Analysis; SE: standard error.
comparable accuracy when identifying auditory-only or AV noise-vocoded words. None of the analyses yielded significant effects that involved the factor Group, in line with previous studies that found comparable performance in (audiovisual) speech perception tasks for dyslexic and typical readers (e.g., see Baart et al., 2012 for adult data; Gijbels et al., 2023, for data obtained in children). Presumably, lip-read information presented in blocks of intermittent AV trials informed the listeners about sound identity and drove learning, which produced adaptation to auditory-only NVS items that was characterized by an overall increase in accuracy across auditory test blocks. This increase is unlikely to originate from accumulating experience with the auditory noise-vocoded speech signal as the experiment progressed. In a prior study with typical readers (Pourhashemi et al., 2022), we presented the exact same words in a highly similar paradigm (i.e., the only differences with the current study were that participants typed in their responses rather than providing it verbally, and the NVS comprised four rather than six channels), and observed a negligible increase (∼3%) in auditory accuracy when participants saw a static image of the speaker’s face during audiovisual training blocks. In contrast, when participants saw dynamic lip-read videos during audiovisual training, auditory-only performance increased with ∼20% (compared to ∼16% in current study). Given that we observed no statistically significant difference between the dyslexic and comparison groups, our data suggest that lip-read-driven adaptation — at a whole-word level — is comparable for dyslexic and typical readers. However, this null effect needs to be interpreted with caution, as it does not necessarily imply that groups are indeed comparable. Therefore, further study is needed to investigate the role of lip-read-driven adaptation in DD in more detail.

As noted, it has been suggested that the effect of lip-read information on speech perception is not as pronounced for dyslexic readers as it is compared to typical readers (e.g., de Gelder and Vroomen, 1998), but at the same time, lip-read-induced phonetic recalibration seems to be intact in readers with DD (Baart et al., 2012; Keetels et al., 2018). In fact, it has even been argued that readers with DD may rely more on lip-read information than typical readers do, in order to compensate for dyslexia-related atypicalities in auditory processing (Pekkola et al., 2006; Schaadt et al., 2016). This aligns with the work by Francisco et al. (2017) who observed that adults with DD who scored lower on phonological awareness were more accurate on a silent lip-read task. Although the current experiment does not allow us to confirm this ‘compensation hypothesis’, at the very least, we observed no evidence that whole-word adaptation to auditory NVS driven by lip-read speech was modulated by dyslexia.

To reconcile the seemingly contrasting findings across the literature, it is of importance to note that DD’s use of lip-read information may be related...
to task demands. In a recent study with 9-to-13-year-olds (Galazka et al., 2021), participants’ eye movements were tracked when listening to auditory sentences or single nonwords while the corresponding talking face was presented. Although the overall proportion of time spent looking at the mouth was comparable for children with and without DD, better readers in the DD group looked longer at the mouth in the nonword condition — which was assumed to be more phonologically demanding than the sentence condition. Related to this, Megnin-Viggars and Goswami (2013) presented adults with four- and 16-channel noise-vocoded speech and demonstrated that, at the group level, supporting lip-read information presented simultaneously with the sound yielded a comparable perceptual advantage for typically reading adults and adults with DD. This was also observed when the visual signal comprised a mosaic version of the (inverted) talking face in which detailed lip-read information was lost, but the visual gain induced by this low-frequency visual information was smaller for participants with poorer auditory-processing capacity. Thus, under challenging listening conditions, participants with less severe reading problems look more at the mouth (Galazka et al., 2021) and participants with less severe auditory-processing issues show more visual gain from low-level dynamic visual cues (Megnin-Viggars and Goswami, 2013). Given that the dyslexic readers in Baart et al. (2012), Keetels et al. (2018) and the current study were all university students capable of performing at an academic level, it is conceivable that their dyslexia symptoms were also relatively mild (although we have no access to their diagnosis). If so, the fact that they relied quite heavily on lip-read information in response to the challenging listening conditions aligns with the work by Galazka et al. (2021).

Van Laarhoven et al. (2018) also provided university students (from the same population as in the current study) with a challenging listening environment by presenting speech in background noise. However, although the items were the same as in the current study (but produced by a different speaker), van Laarhoven et al. (2018) did observe that the effect of lip-read information was less pronounced for the dyslexic readers rather than the comparison group (but see Gijbels et al., 2023 for contrasting findings in 6.5–15-year-olds). Possibly, the contrast between the findings from van Laarhoven et al. (2018) and the current study are related to the fact that in noise-masked speech — unlike in phonetically ambiguous speech or noise-vocoded speech — ambiguities in the signal are not an intrinsic component of the speech signal itself, but arise from an external source and the perceptual system basically needs to reconcile three signals: speech, background noise, and lip-read information. Perhaps this type of stimulus draws rather heavily on the attentional system — which is argued to be impaired in dyslexia (e.g., Facoetti et al., 2003; Harrar et al., 2014; Krause, 2015; Lallier et al., 2010) — leaving insufficient resources for lip-read speech to boost spoken-word recognition.
This study provides valuable insights regarding the supporting role of lip-read information for dyslexic and typical readers in adaptation to noise-vocoded speech. However, there are also limitations to the current work. Most notably, we have no access to our participants’ actual DD diagnosis, the severity of their symptoms, and how dyslexia surfaced exactly. This implies that, at this stage, we cannot generalize our findings across the entire population of individuals with dyslexia. Especially relevant is the question whether dyslexia is related to decoding, comprehension (with intact decoding), or a combination of both. Given that learning effects obtained with noise-vocoded speech are most likely constituted on a pre-lexical rather than lexical level (Hervais-Adelman et al., 2008), and lip-read information is often observed to modulate auditory perception at the level of single phonemes (e.g., Bertelson et al., 2003; McGurk and MacDonald, 1976), it is possible that dyslexic readers with decoding difficulties would benefit most from supporting lip-read information. However, since we do not know the degree to which our dyslexic participants experienced difficulties with decoding and/or comprehension, future work should assess this directly, preferably also by testing accuracy at the phoneme level (which should be relatively difficult for dyslexic readers with decoding difficulties) and sentence (or multiple sentences) level (which should be relatively difficult for dyslexic readers with comprehension difficulties).

The auditory NVS items in the current study were rather difficult to perceive (and accuracy levels reflect those that could be expected for speech in noise). Yet, the absence of a difference between the DD and comparison groups resembles the pattern of lip-read-driven phonetic recalibration observed with (mild) phonetically ambiguous speech (Baart et al., 2012; Keetels et al., 2018, please note that in Baart et al., 2012, the dyslexic readers showed a less-well defined auditory phoneme boundary, but lip-read-driven recalibration was comparable to that in the comparison group). Although future work is clearly needed, this suggests that the critical factor that does or does not allow dyslexic readers to utilize and learn from a lip-read signal is not the level of auditory degradation per se, but whether or not the ambiguities in the signal are intrinsic to the speech signal itself. Variability in idiosyncratic properties of speakers we encounter on a daily basis (regional accents, nonnative pronunciations, particularities in producing certain phonemes) require us to flexibly adjust our system to accommodate for those. If we are to understand a speaker, we need to map these variations onto our existing phonetic categories. This is presumably what happens when we encounter ambiguous speech in combination with lip-read information that disambiguates the sound. As shown by Kraljic et al. (2008), the system indeed learns to accommodate for idiosyncratic properties in the speech signal, but not when these properties can be ascribed to an external, incidental factor (a pen in the speaker’s mouth). Perhaps then, auditory incidental factors that are responsible for sound ambiguities (as is the case for
speech presented in background noise) can also prevent lip-read-driven learning, but only when these tax the listener’s (attentional) system to capacity, as might be the case for dyslexic readers.

Supplementary Material

Supplementary material is available online at:
https://doi.org/10.6084/m9.figshare.25756062

The raw data (anonimized, in.xlsx format) are publicly available for download from https://surfdrive.surf.nl/files/index.php/s/VnGpz5mHk3bmkBB. The unaltered (e.g. non-anonimized).edat files (the E-prime recordings for all individual participants) are available upon request.

Note

1 This declaration is provided after receiving a Dyslexia diagnosis from a registered professional. The full diagnosis report, however, is not included in the declaration document.

References


JASP Team (2020). JASP (Version 0.12) [Computer software]. Available at https://jasp-stats.org/.


