Audiovisual temporal sensitivity in typical and dyslexic adult readers

Ana A. Francisco [1], *Alexandra Jesse* [2], *Margriet A. Groen* [1], *James M. McQueen* [1,3,4]

[1] Radboud University Nijmegen, Behavioural Science Institute, Nijmegen, The Netherlands
[2] Department of Psychology, University of Massachusetts, Amherst, MA, USA
[3] Donders Institute for Brain, Cognition and Behaviour, Nijmegen, The Netherlands
[4] Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands

{a.alvesfrancisco, m.groen, j.mcqueen}@pwo.ru.nl, ajesse@psych.umass.edu

Abstract

Reading is an audiovisual process that requires the learning of systematic links between graphemes and phonemes. It is thus possible that reading impairments reflect an audiovisual processing deficit. In this study, we compared audiovisual processing in adults with developmental dyslexia and adults without reading difficulties. We focused on differences in cross-modal temporal sensitivity both for speech and for non-speech events. When compared to adults without reading difficulties, adults with developmental dyslexia presented a wider temporal window in which unsynchronized speech events were perceived as synchronized. No differences were found between groups for the non-speech events. These results suggests a deficit in dyslexia in the perception of cross-modal temporal synchrony for speech events.

Index Terms: dyslexia, cross-modal temporal sensitivity, speech perception.

1. Introduction

Developmental dyslexia is a common learning disorder. It is characterized by severe difficulties in attaining an adequate reading level, despite normal intelligence and educational opportunities and in the absence of any sensory or neurological impairment [1]. Reading entails a series of complex perceptual and cognitive processes ranging from visual integration to articulation [2]. Accordingly, reading failure could have a wide range of possible causes, any of which could contribute to dyslexia. Over the years, research has suggested that deficient phonological processing is the main underlying cause of reading difficulties [3]. This view has, however, been questioned and the debate has been extended to other processing levels (e.g., auditory [4], visual [5], attentional [6], and cerebellar [7]). Considering that reading is a multimodal process, it is plausible that reading difficulties are, at least partially, explained by an underlying problem with learning, retaining, and accessing audiovisual associations. Though the extent and nature of the relationship between audiovisual processing and reading is still largely unknown, the few existing studies on this topic seem to point toward differences in multisensory processing among individuals with and without reading or other language-learning impairments [8]-[10].

Differences in the efficiency of audiovisual processing might emerge from individual differences in the size of the time window during which audiovisual processing takes place [11]. The perceptual system is tolerant of temporal asynchronies: multisensory signals are still perceived as single temporal events, even when their auditory and visual component are presented with a delay by several hundred milliseconds, at least when the visual component leads [12]. For instance, adults without reading difficulties perceived non-speech audiovisual signals as synchronous between an audio lead of 80 milliseconds and a visual lead of 160 milliseconds [13]. Similarly, they perceived speech stimuli as synchronous between an audio lead of 35 milliseconds and a video lead of 225 milliseconds [14]. Individuals may vary in the size of the time window during which they perceive intersensory synchrony. This variability may relate to individual and group differences in cross-modal temporal sensitivity [11], [12]. In fact, it has been hypothesized that an expanded time window for audiovisual processing may result in difficulties in processes that are dependent on the rapid and accurate binding of cues from multiple senses – processes such as reading [15]. Expanding the time window within which audiovisual integration takes place might result in inappropriate binding of information from the two modalities and, consequently, in less efficient decoding. Adults with developmental dyslexia have been shown to be less sensitive to cross-modal timing than adults without reading impairment, at least for non-speech events [15].

In the present study, we examined cross-modal temporal sensitivity in adults with developmental dyslexia and in adults without reading difficulties with a simultaneity judgment task. While in previous studies [15] cross-modal temporal sensitivity in dyslexia was only assessed for non-speech stimuli, we assessed it in a simultaneity judgment task which measured processing of both speech and non-speech events.

We predict that adults with dyslexia show less cross-modal temporal sensitivity, that is, a wider time window, than adults without reading difficulties. If these differences emerge for both speech and non-speech events, then this reflects a domain-general deficit in dyslexics. If differences are only observed for speech processing, then this suggests a speech-specific deficit.

2. Methods

2.1. Participants

Twenty typical readers (4 males, $M_{age} = 22$, $SD = 2$) and 20 dyslexic readers (3 males, $M_{age} = 23$, $SD = 3$) participated. All participants were native Dutch speakers and undergraduate students at the Radboud University Nijmegen. Participants' hearing was screened with a standard audiometric test. All participants had pure-tone average hearing thresholds below 30 dB for 250, 500, 1K,

2K, 3K, 4K, 6K, and 8K Hz. Reading and phonological awareness were assessed with tasks from a standardized Dutch reading and writing battery for dyslexia diagnosis (*Gl&schr - Test voor gevorderd Lezen en Schrijven*) [16]. Accuracy and time to complete the task were measured in both tasks. To assess reading, participants were asked to read a text out loud while being audio-recorded. Phonological awareness was assessed with a word reversal task. In this task, participants were asked to decide whether the second nonword that they heard in a pair was the phonological reverse of the first one. Participants received monetary compensation or course credits for taking part in the study.

2.2. Materials

The speech event used was a McGurk stimulus, where participants should perceive /ata/ when listening to the syllable /apa/ while seeing a speaker pronouncing /aka/ [17]. These stimuli were taken from [18]. To create these stimuli, a female native Dutch speaker had been video recorded pronouncing the syllables /apa/ and /aka/. An incongruent McGurk stimulus was then created by combining the audio portion of an /apa/ token with the video of an /aka/ token. The release of the /p/ was aligned with the release of the original /k/. White noise was added at -16 dB SNR.

The non-speech event shows a woman clapping her hands. The hands were kept above the head, so that just hands and wrists were visible. Non-speech stimuli were not embedded in noise. The original clapping stimulus was chosen so that its natural asynchrony was identical to that of the original speech stimulus, that is, the time between the start of visual preparation (the beginning of the mouth or the hand movement) and release (the onset of the first phoneme or of the clapping sound) was similar.

All stimuli (speech and non-speech) were recorded with a Sony DCR-HC1000E camera. The videos were digitized as uncompressed 400×320 .avi files in PAL format. The audio was simultaneously recorded at 44.1 kHz. Videos were post-processed and manipulated using Adobe Premiere Elements 11.0 (Adobe Systems, Mountain View, CA) and Praat [19] in order to create the SOAs. The SOAs were created by moving segments, in 40 millisecond increments (one frame), from the beginning of the original file to its end (for the audio leads) and from the end of the original file to its beginning (for the audio lags). This process resulted in the creation of 46 stimuli (23 speech stimuli and 23 non-speech stimuli) with SOAs ranging from (-)440 milliseconds (=auditory lead) to (+)440 milliseconds (=visual lead).

2.3. Procedure and design

Participants were tested during two sessions. The hearing screening, reading-related tasks and the non-speech simultaneity judgment task were performed during the first session. The speech simultaneity judgment task was completed during the second session, together with some additional cognitive measures.

Cross-modal temporal sensitivity for audiovisual speech and non-speech events was assessed with a simultaneity judgement task. Participants saw a speaker say /aka/ and heard /apa/ in the speech condition, and heard and saw someone clap their hands in the non-speech condition. Participants were asked to indicate as fast and as accurately as possible by button press whether the auditory and the visual component of an audiovisual event was presented in synchrony or not. A total of eight blocks were presented for each of the tasks. A block consisted of 23 stimuli (22 asynchronous and one synchronous) presented in random order. A total of 184 trials was thus presented per task. Each trial followed the same presentation: (1) a 50 millisecond black screen; (2) a fixation cross presented for 250 milliseconds; (3) a 200 millisecond black screen; (4) the stimulus presentation. The videos were played in the center of the screen. Each video lasted 2000 milliseconds and was always played completely. After the video offset, the response options were presented and the participants were asked to make a decision about the synchrony of the video they had just seen by pressing one of two response buttons. If a response was not given in five seconds, the next video was presented. No feedback was given. The experimental blocks were preceded by a practice block used to familiarize the participant with the procedure.

Presentation software (Version 16.5, www. neurobs.com) was used to present the experimental tasks on a CRT monitor Iiyama vision master pro451. The refresh rate of the monitor was set to a multiple of the videos' frame rate to guarantee a temporally accurate presentation of the stimuli. The audio was presented via Sennheisser headphones.

3. Results

Adults with developmental dyslexia performed significantly worse than controls on the reading task, both in terms of accuracy (computed as the number of words read incorrectly) (dyslexic readers $M = 7.90$, $SD = 2.02$; typical readers $M = 1.20$, $SD = 1.11$, $t(29.406) = -12.996$, $p < .0001$) and of time required to complete the task (dyslexic readers $M = 335$, $SD = 34.57$; typical readers $M = 210.85$, $SD = 22.05$, $t(32.264) = -13.540$, $p < .0001$). On the phonological awareness task, dyslexic readers were as accurate as controls (computed as the number of hits) (dyslexic readers $M = 17.55$, $SD = 2.06$; typical readers $M = 17.45$, $SD = 2.06$, $t(38) = -.153$, $p = .879$), but took significantly longer to complete the task (dyslexic readers $M = 159.20$, $SD = 40.93$; typical readers $M = 92.50$, $SD = 7.79$, $t(20.375) = -7.160$, $p < .0001$).

To assess cross-modal temporal sensitivity, the percentage of synchronous responses was determined for each participant at each SOA, separately for the speech and non-speech condition. Figures 1 and 2 show the average results for adults with and without dyslexia for the speech and non-speech conditions, respectively. These figures suggest that for speech stimuli, adults with dyslexia and controls are equally sensitive to auditory leads, but dyslexics are less sensitive than controls to visual leads. There seem to be no reading group differences for the non-speech condition. Furthermore, the window of perceived simultaneity seems to be wider for speech than for non-speech events for both groups.
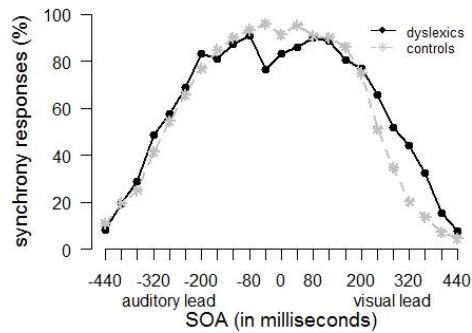
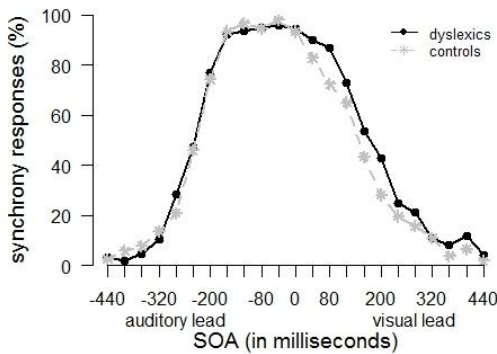Figure 1. Simultaneity responses by SOA and group for the speech events.

Figure 2. Simultaneity responses by SOA and group for the non-speech events.

To test these observations, Gaussian curves were fitted to the data of individual participants, separately for the speech and non-speech condition (see e.g. [12]). These fits were done using the nls function in the stats package in R (version 3.0.3, [20]). This fitting procedure allowed us to estimate the center and the width of the synchrony windows and to test for differences between groups on these two measures. The data from two dyslexic and one control participant had to be removed since for them the width of the window could not be estimated (they never perceived asynchrony for at least one type of lead).

The center of the fitted Gaussian corresponds to the point of subjective simultaneity (PSS), i.e., the average time by which one stimulus has to lead the other to create maximal certainty about stimulus synchrony [21]. A Welch's t-test for unequal variances revealed no differences were observed between groups in this measure (speech events: $t(34.994) = 1.348$, $p = .186$; non-speech events: $t(34.949) = 0.849$, $p = .401$).

The width of the synchrony window was estimated using the full width at half maximum (FWHM) of the fitted Gaussian curve. Participants with dyslexia had a significantly wider window of perceived simultaneity for speech events than participants with typical reading, $(t(34.919) = -2.134$, $p = .039)$. No difference between groups was found, however, for non-speech events $(t(30.726) = -0.942, p = .353)$.

Based on the center and the width, the auditory-lead and the visual-lead end points of the window were also calculated. Table 1 presents the mean estimates of PSS, FWHM, and auditory- and visual-lead threshold, calculated after the individual fits.

Table 1. Curve fits for speech and non-speech events by group[1]

	Group	PSS	FWHM	A-lead	V-lead
S	Dys	-2 (47)	595 (122)	-597 (124)	594 (137)
	Con	-23 (49)	510 (122)	-533 (150)	487 (111)
NS	Dys	-30 (49)	391 (65)	-420 (61)	361 (97)
	Con	-44 (54)	364 (102)	-408 (87)	320 (137)

[1] Means (standard deviations). S = speech events; NS = non-speech events; PSS = point of subjective simultaneity / curve center; FWHM = full width at half maximum / width of synchrony window; A-lead = auditory-lead thresholds; V-lead = visual-lead thresholds.

4. Discussion

The main aim of the present study was to test whether adult dyslexic readers differ from adults without reading impairment in terms of their cross-modal temporal sensitivity. Furthermore, we aimed to test whether any differences in cross-modal temporal sensitivity in dyslexics are limited to perceiving the synchrony in audiovisual speech or whether they also affect synchrony perception of non-speech and are hence domain general. We predicted to observe an expanded temporal window in dyslexic readers.

Two main results were found: first, adults with developmental dyslexia showed a wider temporal window than adult typical readers; second, this was true only for speech events. The observation of a wider temporal window for dyslexic readers confirms our first prediction and is in line with previous evidence of differences in audiovisual speech processing between adults with and without reading impairment [9], [10]. It reaffirms the idea that a multisensory integration deficit may be associated with reading impairment. The plausibility of an audiovisual deficit as a proximal cause for reading deficits in dyslexia has been previously advanced [22]. Reading acquisition requires the development of audiovisual grapheme-phoneme associations, that is, it implies the learning of the systematic links between letters and speech sounds and the automatization of those associations [23]. Previous research showed that these processes require narrow time windows in order to occur adequately [24]. Hence, if grapheme-phoneme associations are established in wider temporal windows, that may foreshadow the development of less appropriate letter-sound correspondences, which, in turn, could hamper reading acquisition. That is, the wide temporal window observed in dyslexic readers might cause, by impairing initial grapheme-phoneme associations, reductions in both the speed and accuracy with which printed representations are decoded [15]. Since we have observed that there is a wider temporal window in adult dyslexic readers, it may be added that the impaired audiovisual processing present in dyslexia seems to represent a persistent condition, rather than a transient developmental lag associated with the beginning of reading acquisition.

Our finding of a wider window in the dyslexic readers in the speech condition, but not in the non-speech condition has a less straightforward explanation. On the one hand, it is consistent with a) studies showing that dyslexic readers perform more poorly at the linguistic than at the non-linguistic level [25] and b) the assumption that dyslexics' temporal processing deficits may be exacerbated at the

level of language [26]. Synchrony judgements might also be generally easier to make for non-speech events (hands clapping) than for speech events. This could be because the onset of the closure and its release are visually more salient in non-speech stimuli, such as clapping, than in speech. It could also be due to a difference in complexity: it has been shown that cross-modal temporal discrimination performance is better for audiovisual stimuli of lower complexity [27] and clapping appears to be a less complex stimulus than speech. This may have made synchrony judgements in the non-speech version of the task easy for both typical and dyslexic readers, thus eliminating group differences. On the other hand, this lack of a group difference for the synchrony judgements of non-speech stimuli contradicts the results of previous studies [15], which report lessened cross-modal temporal sensitivity for non-speech events in dyslexic adults. Methodological differences may underlie this lack of consistency: both the stimuli and the tasks employed are distinct between studies. Whereas we used a clapping stimulus for which perceivers assume unity, the earlier study [15] used materials (circles and tones) that did so to a lesser extent. Furthermore, while we used synchrony judgement tasks, they [15] used a temporal order judgment task. It has been shown that these two tasks are not equivalent, the former one being preferable when the primary interest is in the perceived audiovisual synchrony. Moreover, temporal order judgment tasks seem to be, when compared to synchrony judgement tasks, more susceptible to individual response strategies and less prone to reflect an accurate point of subjective simultaneity [28]. Even though further clarification might be needed regarding the latter methodological points, our results suggest that dyslexics might be no different from typical readers in audiovisual processing of non-speech stimuli.

In summary, our results suggest the presence of a cross-modal deficit for the temporal processing of speech but not for non-speech in adult dyslexic readers. Whether or not this deficit is indeed language specific or occurs more generally under certain processing situations has to be determined. Future research might address this question by systematically varying properties of speech and non-speech stimuli, such as their complexity or visual salience. The current results nevertheless suggest that it is possible that a cross-modal deficit might lead to reading impairment through weakened grapheme-phoneme associations.

5. Conclusions

Since reading is an audiovisual process that requires the learning of systematic links between graphemes and phonemes, it is possible to hypothesize that reading impairments might reflect an audiovisual deficit. Our results are in accordance with such a hypothesis, indicating that adults with developmental dyslexia present less cross-modal temporal sensitivity for speech events than adults without reading impairment.

6. Acknowledgements

7. References

[1] Lyon, R. and Shaywitz, A., "A definition of dyslexia", *Annals of Dyslexia*, 53, 1-14, 2003.

[2] Indefrey, P. and Levelt, W., "The spatial and temporal signatures of word production components", *Cognition*, 92, 101-144, 2004.

[3] Snowling, M. J., *Dyslexia*, Oxford, UK: Blackwell Publishers, 2000.

[4] Tallal, P., "Language disabilities in children: a perceptual or linguistic deficit?", *Journal of Pediatric Psychology*, 5(2), 127-140, 1980.

[5] Livingstone, M. S., Rosen, G. D., Drislane, F. W. and Galaburda, A. M., "Physiological and anatomical evidence for a magnocellular defect in developmental dyslexia", *Proceedings of the National Academy of Sciences*, 88, 7943-7947, 1991.

[6] Facoetti, A. and Molteni, M., "The gradient of visual attention in developmental dyslexia", *Neuropsychologia*, 39 (4), 352-357, 2001.

[7] Nicolson I. and Fawcett, J., "Automaticity: A new framework for dyslexia research?", *Cognition*, 35 (2), 159-182, 1990.

[8] Blau, V., Van Atteveldt, N., Ekkebus, M., Goebel, R. and Blomert, L., "Reduced neural integration of letters and speech sounds links phonological and reading deficits in adult dyslexia", *Current Biology*, 19(6), 503–508, 2009.

[9] Hayes, A., Tiippana, K., Nicol, G., Sams, M. and Kraus, N., "Integration of heard and seen speech: a factor in learning disabilities in children", *Neuroscience Letters*, 351(1), 46–50, 2003.

[10] Norrix, L. W., Plante, E. and Vance, R., "Auditory-visual speech integration by adults with and without language-learning disabilities", *Journal of Communication Disorders*, 39, 22–36, 2006.

[11] Grant, W. and Seitz, F., "Measures of auditory-visual integration in nonsense syllables and sentences", *The Journal of the Acoustical Society of America*, 104(4), 2438–50, 1998.

[12] Conrey, B. and Pisoni, D., "Auditory-visual speech perception and synchrony detection for speech and nonspeech signals", *The Journal of the Acoustical Society of America*, 119(6), 4065-4073, 2006.

[13] Lewkowicz, D., "Perception of auditory-visual temporal synchrony in human infants", *Journal of Experimental Psychology – Human Perception and Performance*, 22 (5), 1094-1105, 1996.

[14] Grant, W., Greenberg, S., Poeppel, D. and van Wassenhove, V., "Effects of spectro-temporal asynchrony in auditory and auditory-visual speech processing", *Seminars in Hearing*, 25, 241-255, 2004.

[15] Hairston, W., Burdette, J., Flowers, D., Wood, F. and Wallace, M., "Altered temporal profile of visual-auditory multisensory interactions in dyslexia", *Experimental Brain Research*, 166(3-4), 474-480, 2005.

[16] De Pessemier, P. and Andries, C., *Test voor Gevorderd Lezen en Schrijven*, Antwerpen, Belgium–Apeldoorn, The Netherlands: Garant, 2009.

[17] McGurk, H. and MacDonald, J., "Hearing lips and seeing voices", *Nature*, 264, 746-748, 1976.

[18] Groen, M. A. and Jesse, A., "Audiovisual speech perception in children and adolescents with developmental dyslexia: No deficit with McGurk stimuli", In S. Ouni, F. Berthommier and A. Jesse [Eds], *Proceedings of the International Conference of Audiovisual Speech Processing* (p. 77-80), 2013.

[19] Boersma, P., *Praat, a system for doing phonetics by computer*, Glot Int. 5, 341–345, 2001.

[20] R Core Team, *R: A language and environment for statistical computing*, R Foundation for Statistical Computation, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org/, 2014.

[21] Zampini, M., Shore, D. I. and Spence, C., "Audiovisual prior entry", *Neuroscience Letters*, 381, 217-222, 2005.

[22] Blomert, L., "The neural signature of orthographic–phonological binding in successful and failing reading development", *NeuroImage*, doi:10.1016/j.neuroimage.2010.11.003, 2011.

[23] Treiman, R. and Kessler, B., "Learning to read", In M. G. Gaskell [Ed.], *Oxford Handbook of Psycholinguistics* (pp. 657-666), Oxford, England: Oxford University Press, 2007.

[24] Blomert, L. and Froyen, D., "Multi-sensory learning to read", *International Journal of Psychophysiology*, 77(3), 195-204, 2010.

[25] Meyler, A. and Breznitz, Z., "Visual, auditory and cross-modal processing of linguistic and non-linguistic temporal patterns among adult dyslexic readers", *Dyslexia*, 11, 93-115, 2005.

[26] Tallal, P., Merzenich, M. M., Miller, S. and Jenkins, W., "Language learning impairments: Integrating basic science, technology, and remediation", *Experimental Brain Research*, 123, 210–219, 1998.

[27] Vatakis, A. and Spence, C., "Audiovisual synchrony perception for music, speech, and object actions", *Brain Research*, 1111, 134-142, 2006.

[28] van Eijk, R. L. J., Kohlrausch, A., Juola, J. F. and van de Par, S., "Audiovisual synchrony and temporal order judgments: Effects of experimental method and stimulus type", *Perception & Psychophysics*, 70(6), 955-968, 2008.