

Audiovisual Speech Perception in Children and Adolescents With Developmental Dyslexia: No Deficit With McGurk Stimuli

Margriet A. Groen¹, Alexandra Jesse²

¹Behavioural Science Institute, Radboud University Nijmegen, Nijmegen, The Netherlands

²Department of Psychology, University of Massachusetts, Amherst, MA, USA

m.groen@pwo.ru.nl, ajesse@psych.umass.edu

Abstract

Developmental dyslexia could, at least partially, reflect an underlying problem in forming audiovisual associations, such as between graphemes and phonemes. Some of the few studies testing people with reading difficulties on McGurk stimuli report less sensitivity to visual information, and worse processing of visual-only speech. In this study, we tested Dutch children ($M = 11.0$ years) and adolescents ($M = 13.7$ years) with developmental dyslexia, and age-matched controls. Dyslexics and age-matched controls were similarly able to recognize the nonsense syllables “apa” and “aka” from hearing or seeing a speaker. Most critically, dyslexics and controls showed similar response patterns to McGurk stimuli, consisting of hearing “apa” combined with seeing a speaker say “aka”. Adolescents, however, perceived McGurk stimuli more often as /k/ and somewhat less often as /p/ than children, confirming earlier studies investigating age differences. Both groups did not differ in their number of fusion (/t/) responses. Concluding, audiovisual speech perception does not seem to be impaired in developmental dyslexia, if groups show similar unimodal speech perception.

Index Terms: speech perception, dyslexia, McGurk effect, development

1. Introduction

Learning to read is essentially an audiovisual mapping problem: Beginning readers need to learn and retain how visual graphemes and auditory phonemes map onto each other. During the reading process, it is important for readers at any reading level to retrieve the phonological information associated with given graphemes [1], [2]. Reading difficulties could thus at least partially reflect an underlying problem with learning, retaining, and accessing audiovisual associations.

Developmental dyslexia is diagnosed in cases of a severe delay in the acquisition of reading in the absence of delays in other cognitive abilities [3]. The delay in reading acquisition cannot be explained by inappropriate educational opportunity or gross neurological disorders. Phonological deficits are a hallmark of dyslexia [4] and are characterized by severe difficulties in decoding and manipulating the phonemes of one’s native language [5]. Phonological deficits cannot be explained through sensory deficits, since only some children and adults with dyslexia have impaired auditory and visual perception (e.g., [6–8]). Rather, phonological deficits seem to reflect problems in accessing phonological representations [9–11]. Reading difficulties may thus emerge as a problem of accessing

phonological representations from visual letter information. A recent functional magnetic imaging study has linked phonological processing deficits in a group of adult dyslexics to impaired audiovisual mapping of letters onto speech sounds [12].

The extent and nature of the relationship between audiovisual processing and reading is, however, still largely unknown. One possibility is that dyslexics have a general deficit in processing and combining audiovisual information. Audiovisual perception can be investigated by making use of the audiovisual nature of speech. Listeners typically use information obtained from seeing a speaker talk in order to comprehend what that speaker says (e.g., [13]). The processing of audiovisual speech partly shares the neural circuit involved in the processing of letter-speech sound associations [14]. This provides an opportunity to investigate audiovisual processing in people with reading impairments in an ecologically valid context (i.e., that of audiovisual speech perception), without probing the direct area of difficulty (i.e., reading).

Audiovisual speech perception can be assessed with the McGurk illusion, where, for example, /ta/ is perceived when the auditory syllable /pa/ is presented with an articulating mouth producing /ka/ [15]. A simplified explanation for this fusion percept is that an alveolar /t/ best matches the contradicting place of articulation information provided by the visual velar /k/ and the auditory bilabial /p/. The few studies testing people with reading difficulties on McGurk illusions provided mixed results. In one study dyslexic children (8-14 years) provided fewer fusion responses, but more responses that were influenced by the visual modality [16] but in another they did not perform differently than their age- or reading-matched controls [17]. Reading-impaired children with broader language impairments from two age groups (6-9 and 10-12 years) and their younger (6-9 years) but not their older (10-12 years) controls gave more auditory-based responses to McGurk stimuli than controls [18]. It is, however, unclear whether that resulted in fewer visually-biased (fusion and/or visual) responses in the reading-impaired group. Similarly, dyslexic adolescents [19] and adults with reading impairments with broader language impairments [20] were less influenced by the visual information of McGurk stimuli than controls. Differences in inclusion criteria for participants as well as in test materials likely contributed to these contradicting results. An additional explanation for this mixed set of results could be that a possible deficit in audiovisual speech perception in people with reading difficulties varies with age. In typical development, the influence of visual information on audiovisual speech perception is larger in adults than in children (e.g., [15], [18], [21-22]). Consequently, any deficit or delay in audiovisual speech perception skills might be more pronounced in older children or adolescents with reading

difficulties than in younger children. Indeed, no differences between younger children (aged < 11 years) with reading difficulties and controls were found by [17] and [18], but older children (aged > 11 years) [18], adolescents [19], and adults [20] with reading difficulties showed different response patterns compared to their controls. Finally, a factor that further complicates the interpretation of the results of these few studies is that children [16], [17], adolescents [19], and adults [23], [24] with reading impairments have been found to be less accurate when identifying unimodal visual syllables.

In the current study, we tested audiovisual processing in children and adolescents with developmental dyslexia by using the McGurk effect. To investigate whether audiovisual processing differs in dyslexics compared to controls, we considered all types of responses, that is, fusion responses as well as auditory- and visual-based responses. Furthermore, we also assessed unimodal perception (auditory and visual) in addition to audiovisual perception. Finally, to test the hypothesis that atypical audiovisual speech perception might be more evident later during development, we included both children and adolescents with developmental dyslexia in our sample.

2. Experiment

2.1. Participants

Nine children (seven boys, $M = 10.76$, $SD = 1.31$ years) and ten adolescents (eight boys, $M = 13.87$, $SD = 0.45$ years) with a clinical diagnosis of developmental dyslexia participated in this study. Eight typically developing children (six boys, $M = 11.09$, $SD = 1.34$ years) and 12 adolescents (ten boys, $M = 13.59$, $SD = 0.31$ years) of similar chronological age took part as the control group. All participants obtained a score of at least 35 on the matrix reasoning subtest of a standardized measure of non-verbal cognitive ability ([25] $M = 50$, $SD = 10$) and passed a hearing test (i.e., pure-tone average hearing thresholds were between 30 and 40dB at maximally two frequencies, rest below 30 dB at 250, 500, 1K, 2K, 3K, 4K, 6K, and 8K Hz in both ears). Participants had to obtain a score of 7 or lower on the standardized measures ($M = 10$, $SD = 3$) of word [26] or non-word [27] reading to be included in the dyslexia group; or a score higher than 7 on both tests to be in the control group. Another 11 participants were tested, but their data was excluded from the analyses because they did not pass the hearing screening (7), did not meet the criteria for inclusion in the control group (2) or the dyslexic group (2). Participants received a small present as a thank you for taking part in the study.

2.2. Materials

A female native Dutch speaker was video recorded with a Sony DCR-HC1000E camera as she pronounced the VCV syllables /apa/ and /aka/. Videos showed the face and the neck of the speaker. Videos were digitized as uncompressed 720 × 576 .avi files in PAL format. Audio was recorded at 44.1kHz at the same time with a standalone Sennheiser microphone. Two incongruent McGurk stimuli were created by combining the audio portion of an /apa/ token with the video of an /aka/ token. The release of the /p/ was hereby presented at the same time as the release of the original /k/. The audio and the video portion of the same /apa/ and /aka/ tokens used to create these McGurk stimuli were presented during unimodal auditory and visual trials. All stimuli were presented in -16 dB SNR white noise to increase the

possible influence of visual information on perception in the audiovisual condition. A block consisted of eight unimodal (two tokens x two phonemes x two modalities) and eight audiovisual (two tokens x four presentations) stimuli, presented in random order. A total of 12 blocks was run, consisting of 96 unimodal and 96 McGurk stimuli.

2.3. Procedure and design

Participants were tested in a quiet room in their school. Cognitive and reading tests and the audition screen were performed during the first session. The experimental task was completed during the second session. Presentation software (Version 14.3, www.neurobs.com) was used to present the experimental task on a laptop computer and participants heard the auditory portion via Sennheiser headphones. Participants were asked to report which stimuli they perceived by pressing one of three response buttons, labeled “apa”, “ata”, and “aka”. To the children the task was introduced as a game in which they had to assist alien shuttles during landing by clarifying the noisy communication channel between control tower and pilot through pressing one of the response buttons.

2.4. Results and discussion

As expected, participants with dyslexia performed worse than controls on standardized word reading (dyslexic children $M = 6.22$, $SD = 2.39$; control children $M = 11.75$, $SD = 2.68$; dyslexic adolescents $M = 5.00$, $SD = 2.21$; control adolescents $M = 10.25$, $SD = 1.14$; $F(35) = 63.23$, $p < .001$) and non-word reading (dyslexic children $M = 5.11$, $SD = 1.62$; control children $M = 11.88$, $SD = 2.80$; dyslexic adolescents $M = 5.50$, $SD = 1.84$; control adolescents $M = 11.25$, $SD = 1.82$; $F(35) = 91.69$, $p < .001$) tests. No effects of age group or interactions between age and reading group were found for the reading measures.

Mixed effect models were used to analyze the speech perception data, using the lmer function [28] in the R statistical program [29]. A binomial logit linking function was used for the categorical dependent variables. Systematic step-wise model comparisons established the best-fitting models through likelihood ratio tests. Model comparisons started with a full model, from which non-significant interactions and main effects were gradually removed. The main effects of factors contributing to significant interactions remained in the models. To assess performance on unimodal trials, age group (children, adolescents), reading group (dyslexic, control), and stimulus (/p/, /k/) were evaluated as contrast-coded categorical factors. Age group (children, adolescents), reading group (dyslexic, control), and token (1,2) were assessed as contrast-coded categorical factors to see whether they explained performance on audiovisual trials. One condition of each these categorical fixed factors is mapped onto the intercept of the model. The adjustment for the other level of a variable is then estimated. Adjustments that differ significantly from zero indicate significant effects. Subject was included as random factor in all best-fitting models, allowing for subject-specific adjustments to the regression weights.

Children and adolescents performed similarly on recognizing auditory-only ($\chi^2(1) = 1.02$, $p = .31$) and visual-only syllables ($\beta = -0.61$, $SE = 0.45$, $p = .18$; see Table 1). Dyslexics and their controls also performed similarly on these unimodal trials (auditory-only: $\chi^2(1) = 0.96$, $p = .33$; visual-only: $\chi^2(1) = 0.0006$,

$p = .98$). Age group and reading group did also not interact with each other or with stimulus (all $p > 0.05$). /p/ was more difficult to identify than /k/, when presented auditorily ($\beta = 2.81$, $SE = 0.18$, $p < .001$), but easier to identify, if presented visually ($\beta = -3.5$, $SE = 0.27$, $p < .001$). The latter effect was somewhat larger for children than for adolescents ($\beta = 0.998$, $SE = 0.27$, $p = .07$).

Table 1. Percentage of correct responses for dyslexic (DYS) and control (CON) participants on unimodal trials by stimuli and age group. Standard deviations are given in parentheses.

		Children		Adolescents	
		DYS	CON	DYS	CON
Auditory	/p/	80.73 (18.22)	91.15 (9.82)	82.92 (11.69)	84.72 (10.86)
	/k/	95.83 (8.63)	98.44 (3.10)	97.08 (3.43)	95.14 (7.07)
Visual	/p/	97.92 (4.45)	99.48 (1.47)	95.42 (4.14)	96.53 (5.57)
	/k/	65.1 (17.53)	62.5 (26.54)	62.5 (25.91)	60.76 (24.00)

Table 2. Percentage of responses for dyslexic (DYS) and control (CON) participants on incongruent audiovisual McGurk trials, separated by token and age group. Standard deviations are given in parentheses.

		Children		Adolescents	
		DYS	CON	DYS	CON
Token 1	/p/	27.86 (35.02)	52.08 (43.14)	17.50 (26.75)	16.67 (12.15)
	/k/	27.6 (27.59)	14.32 (33.08)	49.38 (32.81)	54.52 (21.32)
	/t/	44.53 (27.99)	33.59 (38.21)	33.12 (32.31)	28.82 (21.07)
Token 2	/p/	17.19 (33.63)	40.62 (44.64)	16.67 (25.8)	11.63 (11.08)
	/k/	7.29 (15.02)	11.72 (32.31)	42.08 (32.67)	42.19 (29.85)
	/t/	75.52 (33.85)	47.66 (46.08)	41.25 (31.27)	46.18 (32.01)

Table 2 provides the response patterns for audiovisual McGurk trials, separated by token and group. Overall, both tokens elicited a substantial number of fusion responses (token 1 $M = 34.27\%$, $SD = 29.04\%$; token 2 $M = 51.37\%$, $SD = 36.42\%$). Performance on incongruent audiovisual trials was first assessed by coding responses as auditory-based responses (i.e., /p/ responses) versus visually-biased responses. Visually-biased responses consisted of visual-based /k/-responses as well as

fusion responses (/t/). To examine further whether the type of visually-biased responses differed across groups, we also regrouped responses as visual-based responses (/k/) versus all other types of response and then as fusion responses versus all other types of responses.

Most critically, none of the analyses showed a difference in performance between dyslexics and their age-matched controls (all $p > 0.05$). Reading group did also not interact with any other factor (all $p > 0.05$). The analyses showed, however, a difference between age groups. Adolescents gave overall more visual-based /k/ responses than children ($\beta = 3.15$, $SE = 0.8$, $p < .0001$) and somewhat fewer auditory-based responses ($\beta = -1.43$, $SE = 0.84$, $p = .09$). Children and adolescents did, however, not differ in the number of fusion responses ($\beta = -0.52$, $SE = 0.89$, $p = .56$). The analyses also showed that token 2 elicited fewer auditory-based responses ($\beta = -0.83$, $SE = 0.12$, $p < .0001$) and fewer visually-based responses ($\beta = -1.23$, $SE = 0.13$, $p < .0001$) than token 1, but more fusion responses ($\beta = 1.37$, $SE = 0.10$, $p < .0001$). These token effects were always larger for children than for adolescents (auditory: $\beta = 1.1$, $SE = 0.24$, $p < .0001$; visual: $\beta = 1.34$, $SE = 0.26$, $p < .0001$; fusion: $\beta = -1.10$, $SE = 0.20$, $p < .0001$).

3. Discussion

We tested audiovisual processing of speech in Dutch participants with developmental dyslexia by using the McGurk effect. Unlike previous work, we considered all types of responses and also assessed unimodal perception (auditory and visual). Children and adolescents with developmental dyslexia were tested, to also evaluate the hypothesis that atypical audiovisual speech perception might be more evident later during development.

We did not find any group differences in unimodal auditory or visual speech perception. Dyslexics and their age-matched controls were similarly able to recognize the nonsense syllables “apa” and “aka” from hearing and from seeing a speaker. This finding is not in line with results from previous studies that found that children [14-15], adolescents [19] and adults [21-22] with reading impairments were, compared to their controls, less accurate at identifying unimodal visual syllables.

The fact that in our study dyslexics and their controls showed similar auditory and visual processing allowed us to investigate whether these groups differ in their audiovisual processing above and beyond differences in unimodal processing. Being equated on their processing of auditory and visual speech, dyslexics and controls also showed similar response patterns to McGurk stimuli, consisting of an auditorily presented “apa” combined with a visually presented “aka”. At the group level, dyslexics experienced similar proportions of fusion illusions, and also similar proportions of auditory-based and visual-based percepts as their age-matched controls. As can be seen from the large standard deviations in Table 2, however, individual variation in response to McGurk stimuli was considerable in all groups and merits further investigation.

In typical development, the influence of visual information on audiovisual speech perception is larger in adults than in children (e.g., [15], [18], [21-22]), potentially leading to an age-dependent deficit in dyslexia. As such we had expected to find group differences for adolescents, but not for children, with and without developmental dyslexia in our sample. We did, though, not find any differences in audiovisual speech perception

between dyslexics and controls. Our results confirm, however, earlier studies investigating age differences in that adolescents overall perceived McGurk stimuli more often as /k/ and somewhat less often as /p/ than children. Age groups did not differ in their number of fusion (/t/) responses.

4. Conclusions

Forming associations between letters and speech is an imported step in learning to read and developmental dyslexia might, at least partially, reflect a deficit in acquiring, retaining, and accessing audiovisual associations [12]. Dyslexia could hence, at least partly, reflect an audiovisual processing problem. Our study suggests, however, that audiovisual speech perception, as assessed with McGurk stimuli, does not seem to be impaired in developmental dyslexia, when dyslexics show similar skills to those of age-matched controls in the processing of speech from hearing and seeing a speaker alone. The sample size of the current study is modest and only processing of incongruent audiovisual syllables was evaluated. Possibly, acquiring, retaining, and accessing letter-sound associations involves mechanisms different from those used when processing audiovisual speech. Before we can reach that conclusion, however, a more extensive investigation of individual differences in audiovisual speech perception using both incongruent and congruent stimuli, and their associations with performance on reading and reading-related skills, is needed.

5. Acknowledgements

This work was supported by a postdoctoral fellowship award from the Behavioural Science Institute to MG. The authors wish to thank Jennifer Snijder and Judith Pustjens for assistance with recruiting and testing participants, school staff of ‘Patricius’, ‘De Violier’, ‘Maria Goretti’ and ‘Isendoorn College’ for their support and all participants for their cooperation.

6. References

- [1] Frost, R., “Toward a Strong Phonological Theory of Visual Word Recognition: True Issues and False Trails,” *Psychological Bulletin*, 123(1):71–99, 1998.
- [2] Rastle, K. and M. Brysbaert, “Masked phonological priming effects in English: Are they real? Do they matter?,” *Cognitive Psychology*, 53(2):97–145, 2006.
- [3] Snowling, M. J., *Dyslexia*. Oxford, UK: Blackwell Publishers, 2000.
- [4] Ramus, F., “Theories of developmental dyslexia: insights from a multiple case study of dyslexic adults,” *Brain*, 126(4):841–865, 2003.
- [5] Shaywitz, S. E. and B. A. Shaywitz, “Dyslexia (specific reading disability),” *Biological Psychiatry*, 57(11):1301–1309, 2005.
- [6] Tallal, P., “Rapid auditory processing in normal and disordered language development,” *Journal of Speech & Hearing Research*, 19(3):561–71, 1976.
- [7] Ramus, F., “Developmental dyslexia: specific phonological deficit or general sensorimotor dysfunction?,” *Current Opinion in Neurobiology*, 13(2):212–218, 2003.
- [8] Rosen, S., “Auditory processing in dyslexia and specific language impairment: is there a deficit? What is its nature? Does it explain anything?,” *Journal of Phonetics*, 31(3–4):509–527, 2003.
- [9] Shankweiler, D. and S. Crain, “Language mechanisms and reading disorder: A modular approach,” *The Onset of literacy: cognitive processes in reading acquisition*, 24:139–168, 1986.
- [10] Hulme, C. and M. J. Snowling, “Deficits in output phonology: an explanation of reading failure?,” *Cognitive Neuropsychology*, 9(1):47–72, 1992.
- [11] Ramus, F. and G. Szenkovits, “What phonological deficit?,” *The Quarterly Journal of Experimental Psychology*, 61(1):129–141, 2008.
- [12] Blau, V., N. van Atteveldt, M. Ekkebus, R. Goebel, and L. Blomert, “Reduced neural integration of letters and speech sounds links phonological and reading deficits in adult dyslexia,” *Current Biology*, 19(6):503–508, 2009.
- [13] Massaro, D. W. and Jesse A., “Chapter 2: Audiovisual speech perception and word recognition,” in G. Gaskell [Ed], *The Oxford Handbook of Psycholinguistics*, 19–35, Oxford University Press, 2007.
- [14] Blomert, L., “The neural signature of orthographic–phonological binding in successful and failing reading development,” *NeuroImage*, 57:695–708, 2011.
- [15] McGurk, H. and J. MacDonald, “Hearing lips and seeing voices,” *Nature*, 264:746–748, 1976.
- [16] Hayes, E. A., K. Tiippana, T. G. Nicol, M. Sams, and N. Kraus, “Integration of heard and seen speech: a factor in learning disabilities in children,” *Neuroscience Letters*, 351(1):46–50, 2003.
- [17] Bastien-Toniazzo, M., A. Stroumza, and C. Cavé, “Audio-Visual Perception and Integration in Developmental Dyslexia: An Exploratory Study Using the McGurk Effect,” *Current Psychology Letters. Behaviour, Brain & Cognition*, 25(3):2-15, 2010.
- [18] Boliek, C., C. Keintz, L. Norrix, and J. Obrzut, “Auditory-visual perception of speech in children with learning disabilities: The McGurk effect,” *Canadian Journal of Speech-Language Pathology and Audiology*, 34(2):124–131, 2010.
- [19] Campbell, R., A. Whittingham, U. Frith, D. W. Massaro, and M. M. Cohen, “Audiovisual Speech Perception in Dyslexics: Impaired Unimodal Perception but no Audiovisual Integration Deficit” *Proceedings of the International Conference on Auditory-Visual Speech Processing*, 85-88, 1997.
- [20] Norrix, L. W., E. Plante, and R. Vance, “Auditory–visual speech integration by adults with and without language-learning disabilities,” *Journal of Communication Disorders*, 39(1):22–36, 2006.
- [21] Massaro, D. W., “Children’s Perception of Visual and Auditory Speech,” *Child Development*, 55(5):1777–1788, 1984.
- [22] Dupont, S., J. Aubin, and L. Ménard, “A study of the McGurk effect in 4 and 5-year-old French Canadian children,” *ZAS Papers in Linguistics*, 40:1–17, 2005.
- [23] De Gelder, B. and J. Vroomen, “Impaired speech perception in poor readers: Evidence from hearing and speech reading,” *Brain & Language*, 64:269–281, 1998.
- [24] Ramirez, J. and V. Mann, “Using auditory-visual speech to probe the basis of noise-impaired consonant–vowel perception in dyslexia and auditory neuropathy,” *The Journal of the Acoustical Society of America*, 118(2):1122–1133, 2005.
- [25] Wechsler, D. and J. A. Naglieri, *Wechsler Nonverbal Scale of Ability, Dutch Edit., Pearson Assessment*, 2006.
- [26] Brus, B. T. and M. J. M. Voeten, *Een-Minut-Test*. Pearson, 1999.
- [27] Van den Bos, K. P., H. C. L. Spelberg, A. J. M. Scheepstra, and J. R. De Vries, *Klepel*. Pearson, 1994.
- [28] Bates, D. and D. Sarkar, “Lme4: Linear mixed-effects models using s4 classes.” 2007.
- [29] R Development Core Team, “R: A language and environment for statistical computing.” Vienna, 2007.