ELSEVIER

Neuroscience Letters 351 (2003) 46–50

Neuroscience Letters

www.elsevier.com/locate/neulet

Integration of heard and seen speech: a factor in learning disabilities in children

Erin A. Hayes[a,*], Kaisa Tiippana[b], Trent G. Nicol[a], Mikko Sams[b], Nina Kraus[a,c,d]

[a]*Department of Communication Sciences, Northwestern University, Evanston, IL, USA*

[b]*Laboratory of Computational Engineering, Helsinki University of Technology, P.O. Box 9203, Helsinki, 02015 HUT, Finland*

[c]*Departments of Neurobiology and Physiology, Northwestern University, Evanston, IL, USA*

[d]*Department of Otolaryngology, Northwestern University, Evanston, IL, USA*

Received 2 June 2003; received in revised form 21 July 2003; accepted 25 July 2003

Abstract

Normal-learning children (NL) and children with learning disabilities (LD) reported their perceptions of unisensory (auditory or visual), concordant audiovisual (e.g. visual /apa/ and auditory /apa/) and conflicting (e.g. visual /aka/ and auditory /apa/) speech stimuli in quiet and noise (0 dB and − 12 dB signal-to-noise ratio, SNR). In normal populations, watching such conflicting combinations typically changes auditory percepts (‘McGurk effect’). NL and LD children identified unisensory auditory and congruent audiovisual stimuli similarly in all conditions. Despite being less accurate identifying unisensory visual stimuli, LD children were more likely than NL children to report hearing only the visual component of incongruent audiovisual stimuli at − 12 dB SNR. Furthermore, LD children with brainstem timing deficits demonstrated a distinctive pattern of audiovisual perception. The results suggest that the perception of simultaneous auditory and visual speech differs between NL and LD children, perhaps reflecting variations in neural processing underlying multisensory integration.

Keywords: Audiovisual integration; McGurk effect; Learning disabilities; Dyslexia; Speech perception; Auditory brainstem response; Visual perception; Auditory perception

Learning disabilities that are manifested as reading and spelling deficits have been attributed to deficits in the perception and neurophysiologic encoding of both auditory and visual stimuli [8,9]. More specifically, these studies have focused on auditory speech perception impairments [4], deficits in the magnocellular pathway of visual processing [3] and difficulties processing rapidly presented stimuli in any sensory modality [14]. Recent imaging studies have examined neural activity in multisensory cortical regions in normal and learning disabled readers and have discovered altered patterns of connectivity among primary sensory and multisensory processing areas in learning disabled subjects [12,13]. In these individuals, reading deficits may be attributable to deficits integrating multisensory information.

One perceptual illusion that reflects multisensory integration is the McGurk effect, which occurs when people see and hear a talker producing incongruent speech segments [11]. For example, a classic McGurk effect occurs when an auditory /pa/ and a visual /ka/ are presented together. People often hear /ta/, a fusion of the auditory and visual stimuli. When auditory /ka/ is dubbed onto visual /pa/, a typical auditory percept is /pka/ or /kpa/, combinations of the auditory and visual stimuli. Other possible percepts are /pa/, the visual component, or /ka/, the auditory component. Occasionally they may hear a phoneme that fails to incorporate the acoustic or visual aspects of either stimulus; these are categorized as other responses.

In this study, auditory, visual and audiovisual speech perception were examined in normal learning and learning disabled children. Subjects were between the ages of 8 and 14, had normal binaural hearing thresholds (≤ 15 dB HL for octaves 500–8000 Hz), had normal binocular near vision (better than 20/40), were native speakers of American English and had normal mental ability (IQ ≥ 85 on the *Brief Cognitive Scale*, Woodcock-Johnson (WJ) [18]). Standardized measures of cognition and academic achievement were administered (see Table 1). *Cross Out* is a measure of visual processing, including visual scanning and attention [19]. *Auditory Processing* is a composite score derived from *Incomplete Words* and *Sound Blending* [19]. *Listening Comprehension* [19] and *Memory for Words* [19] measure

* Corresponding author. Northwestern University, 2240 N. Campus Drive, Evanston, IL 60208, USA. Tel.: +1-847-491-2465; fax: +1-847-491-2523.

E-mail address: e-hayes@northwestern.edu (E.A. Hayes).

0304-3940/03/$ - see front matter © 2003 Elsevier Ireland ltd. All rights reserved.
doi:10.1016/S0304-3940(03)00971-6

auditory comprehension and memory, respectively. *Reading* and *Spelling* (Wide Range Achievement Test-3 [17]) assess single word reading and single word spelling, and *Word Attack* (WJ-R) [19] assesses nonsense word reading. To evaluate fundamental acoustic processing, subjects listened to speech sounds along a synthetic /da-ga/ continuum. The third formant onset frequency changed in 10 Hz steps from 2580 Hz (/da/) to 2180 Hz (/ga/) and an adaptive tracking procedure was used to determine the difference between the third formant onset frequencies that was necessary for a subject to correctly discriminate between the phonemes 69% of the time [7]. Children without learning disabilities (NL, $n = 10$) scored within normal limits on the psychoeducational test battery. Children with learning impairments (LD, $n = 13$) diagnosed by independent clinicians, exhibited a discrepancy (≥ 12 points) between the *Brief Cognitive Scale* and at least one measure of reading or spelling. The NL and LD groups differed on their scores for *Reading*, *Spelling* and *Word Attack* (Student's *t*-test, $P < 0.001$ for all) but not on tasks that assess auditory discrimination, auditory processing, and visual processing (Table 1).

Stimuli were constructed from a videotape of a female speaker saying /ata/, /apa/ and /aka/ in a clear speaking style. Audio and visual tracks were separated for editing. Delivery of the audio and visual stimuli was controlled by Presentation software (Neurobehavioral Systems, Inc., CA). The auditory stimuli were 600 ms in duration and presented binaurally through Sennheiser HD 580 headphones (Sennheiser Electronics Corp., CT) at 65 dB SPL. Stimuli were presented in three different combinations: unisensory (auditory-only or visual-only); congruent (auditory and visual stimuli the same); and incongruent (auditory and visual stimuli different). Incongruent pairs used in this study were: visual /aka/ + auditory /apa/; visual /apa/ + auditory /aka/; visual /apa/ + auditory /ata/; and visual /ata/ + auditory /apa/. White Gaussian noise was added to the auditory stimuli to create three different signal-to-noise ratios (SNR): quiet (ambient background of 27 dB SPL), 0 dB and −12 dB (white Gaussian noise of 65 and 72 dB SPL).

Stimuli were presented to the children in five blocks of 66 stimuli. During each block, each visual-only stimulus occurred two times and each auditory-only stimulus, congruent audiovisual and incongruent audiovisual stimulus pair occurred two times in each noise condition. All stimuli occurred once in a block before any stimulus was repeated. Children were instructed to watch the face, listen to the voice, and verbally report what they heard; if only a visual signal was presented they were to report what they thought the speaker was saying. The tester sat with her back to the video screen and coded the responses as: k, p, t, kp, kt, pk, pt, tk, tp, or other. Perceptions of /g/, /b/ or /d/, voiced consonants, were coded as the unvoiced consonants /k/, /p/ or /t/, respectively. If the subject was not viewing the screen during a trial, the trial was omitted.

There was no main effect of diagnostic group on the identification of auditory stimuli, nor was there a significant interaction between diagnostic group and noise level or diagnostic group and stimuli (repeated measures analysis of variance, $F_{(1,21)} = 0.02$, $P = 0.88$; $F_{(2,42)} = 1.03$, $P = 0.37$; and $F_{(2,42)} = 2.95$, $P = 0.06$, respectively). However, there was a main effect of diagnostic group on the identification of visual stimuli ($F_{(1,21)} = 5.97$, $P < 0.02$), with a significant interaction between diagnostic group and stimuli ($F_{(2,42)} = 7.90$, $P < 0.001$). Post-hoc tests revealed that NLs identified visual stimuli more accurately than LDs, and that these differences occurred for the visemes /aka/ and /ata/ (Student's *t*-test, $P < 0.001$ and $P < 0.05$, respectively), but not /apa/ (Table 2).

The improvement in perceptual accuracy between

Table 1
Subject characteristics

	NL ($n = 10$)	LD ($n = 13$)
Gender		
M	8	10
F	2	3
Age (years)	11.3 (1.7)	10.7 (1.5)
IQ	124 (14)	114 (12)
Reading*	113 (11)	95 (9)
Word Attack*	118 (9)	94 (13)
Spelling*	114 (15)	91 (8)
Auditory Processing	99 (9)	96 (9)
Memory for Words	111 (16)	101 (11)
Listening Comprehension	123 (17)	118 (17)
Cross Out (visual processing)	112 (12)	106 (12)
Just noticeable difference: /da-ga/ (F3 onset in Hz)	92 (29)	115 (56)
Brainstem response (ms)	7.56 (0.30)	7.73 (0.31)
Normal	7.39 (0.13)	7.49 (0.21)
Delayed	7.89 (0.18)	8.04 (0.45)

Demographic, cognitive, perceptual and neurophysiologic means (and standard deviations) for the two groups. *The NL and LD groups differed significantly on measures of single word reading, spelling, and nonsense word reading. No other measures examined in this study differed significantly between the groups.

unisensory and congruent audiovisual stimuli presentation, defined as enhancement, was calculated as (% correct congruent minus % correct unisensory)/% correct unisensory. For each SNR, enhancement was calculated for the improvement offered by the addition of congruent visual stimuli to auditory stimuli (visual enhancement) and the addition of congruent auditory stimuli to visual stimuli (auditory enhancement). NLs and LDs did not differ in the amount of benefit for either type of enhancement (visual enhancement: $F_{(1,21)} = 0.49$, $P = 0.49$; auditory enhancement: $F_{(1,21)} = 1.52$, $P = 0.23$). Noise increased visual enhancement ($F_{(2,42)} = 67.16$, $P < 0.01$) and decreased auditory enhancement ($F_{(2,42)} = 6.95$, $P < 0.002$). The effects of noise occurred irrespective of diagnostic category (visual enhancement: $F_{(2,42)} = 1.48$, $P = 0.24$; auditory enhancement: $F_{(2,42)} = 0.88$, $P = 0.42$).

Percepts of incongruent audiovisual stimuli were classified as capturing the visual or auditory stimulus, a combination or fusion of the visual and auditory stimuli (e.g. /ata/, /atha/), or reflecting other phonemes (e.g. /asa/, /alma/). Because unisensory perception impacts audiovisual perception, unisensory performances were used as covariates in the analyses of incongruent audiovisual perception.

While both NLs and LDs reported more visual responses with increasing background noise ($F_{(2,42)} = 51.28$, $P < 0.001$; Fig. 1), this pattern was stronger in LDs ($F_{(2,42)} = 3.31$, $P < 0.05$). In the highest background noise, more visual responses were reported by LDs than NLs ($F_{(1,19)} = 4.47$, $P < 0.05$) while more combination/fusion responses were reported by NLs than LDs ($F_{(1,19)} = 5.57$, $P < 0.03$). This suggests that for LDs, the perception of incongruent stimuli was influenced more by the visual stimulus than it was for NLs. No significant differences between the NL and LD groups were observed for incongruent stimulus pairings in easier listening conditions (0 dB SNR and quiet), nor for the reporting of auditory or other responses.

The brainstem response elicited by a speech sound

Table 2
Unisensory identification

	Diagnostic group	
	NL	LD
All phonemes		
Auditory: quiet	78 (10)	82 (15)
Auditory: 0 dB SNR	44 (16)	46 (13)
Auditory: − 12 dB SNR	24 (10)	20 (10)
All visemes*	82 (19)	59 (25)
Visual /aka/*	93 (3)	46 (10)
Visual /ata/*	78 (10)	49 (10)
Visual /apa/	75 (10)	82 (6)

Mean unisensory correct identification percentages with standard deviations in parentheses. * indicates a significant difference between groups ($P < 0.05$). There were no significant differences between groups on phonemes. However, there was a main effect of diagnostic group on the identification of visemes and an interaction of viseme with group.

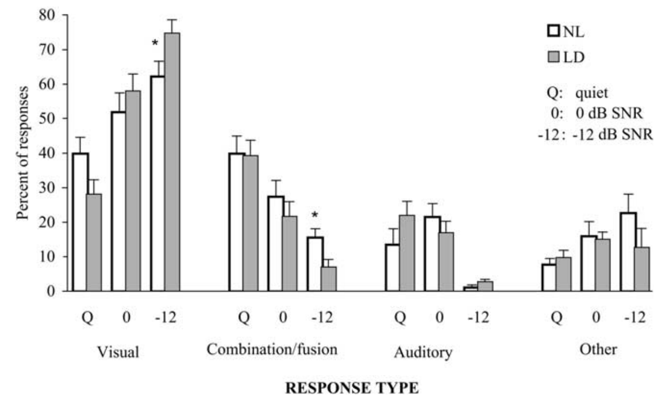

Fig. 1. Response types averaged across the four incongruent stimulus pairs. Presented here are the adjusted means for the response type, calculated with audio-alone and visual-alone scores as covariates; error bars represent one standard error of the mean. At − 12 dB SNR, LD children reported a higher proportion of visual responses to incongruent stimuli. Conversely, combinations/fusions were more likely to be reported by NL children. A similar pattern of results can be seen at 0 dB SNR, but the differences were not statistically significant. In addition, there was an interaction between noise level and diagnostic group such that LDs exhibited a greater increase in perception of the visual component of incongruent stimuli with increased noise level. Auditory and other responses did not differ between groups at any noise level.

previously has been shown to differ between groups of children with learning problems and NL children [5]. It was hypothesized that the bias toward visual responses seen in LDs to incongruent audiovisual stimuli might be due to atypical encoding of speech sounds at the brainstem level. To test this hypothesis, the auditory brainstem response (ABR) was elicited by randomly mixed condensations and rarefactions of a 40 ms speech stimulus (/da/) presented to the right ear through insert earphones at 80 dB SPL with an interstimulus interval of 51 ms. The ABRs were recorded from Cz-to-ipsilateral earlobe, with forehead as ground. The latency of wave A, the largest negativity following wave V in the brainstem complex elicited by /da/, was measured from the neural response. Within the LD group, eight children had normal wave A latencies (within ± 1 standard deviation of the mean [5]; LD-normal) and five exhibited delayed wave A latencies (LD-delayed). Eight of the ten NL children exhibited normal wave A latencies (NL-normal). Two NL children with delayed latencies were omitted from the statistical analysis because of the small sample size (Table 1).

Collapsed across all stimuli, no differences were found among the three groups for the identification of unisensory auditory stimuli nor for the degree of enhancement engendered by congruent stimuli (auditory stimuli: $F_{(2,18)} = 1.29$, $P = 0.30$; auditory enhancement: $F_{(2,18)} = 1.00$, $P = 0.39$; visual enhancement: $F_{(2,18)} = 1.34$, $P = 0.29$). The pattern of visual identification (NL 86%, LD-normal 57%, LD-delayed 61%) was very similar to that of the larger data set but did not reach significance due to a reduced sample size.

The ability to precisely represent speech sounds at the level of the brainstem (reflected by wave A latency) was a factor in the perception of incongruent stimuli at − 12 dB SNR (Fig. 2). Children in the LD-delayed group reported the fewest combination and fusion responses, the NL-normal group reported the most, while the LD-normal group reported an intermediate number ($F_{(2,16)} = 4.29, P < 0.03$). Although the LD-delayed subjects reported fewer combination and fusion responses to incongruent stimuli, there was not a consistent shift toward other response types. Thus, the three groups did not statistically differ in the proportions of visual-only, auditory-only, and other responses that they reported ($F_{(2,16)} = 0.85$, $P = 0.45$; $F_{(2,16)} = 0.61$, $P = 0.55$; $F_{(2,16)} = 0.38$, $P = 0.69$, respectively).

For these groups of NL and LD children, selected to be similar on measures of auditory processing and speech discrimination, it is not surprising that the children performed similarly in the identification of auditory-only stimuli. Also, because the groups were matched on *Cross Out* (a measure of visual attention and processing speed), differences were not expected for the perception of visual-only stimuli. However, LDs were poorer in visual speech reading than NLs, in agreement with a previous finding by de Gelder and colleagues [2]. This may be because the LDs in the present study have a specific deficit in visual speech perception or a deficit in visual perception not assessed by *Cross Out*.

NL and LD children perceived incongruent audiovisual stimuli differently in challenging listening conditions. LDs were more likely than NLs to report hearing the visual component, while NLs were more likely than LDs to report hearing combinations and fusions of the auditory and visual components. These differences could reflect distinct processes at many levels of perception. Auditory and/or visual encoding deficits in the LDs may lead to an increased cognitive demand in processing the neural representations

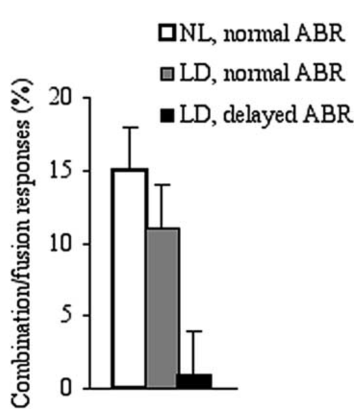

Fig. 2. Combination and fusion responses to incongruent stimuli at − 12 dB SNR for children grouped according to brainstem latency elicited by /da/. Presented here are the adjusted means calculated with audio-alone and visual-alone scores as covariates; error bars represent one standard error of the mean. LD children with delayed latencies reported fewer combination and fusion responses to incongruent audiovisual stimuli than NL children with a normal latency. The frequency of combination/fusion responses in LD children with a latency within normal limits was intermediate between the other two groups.

of the stimuli. Consequently, this shift in resources may affect their perceptions of incongruent audiovisual stimuli in challenging listening conditions.

It is also possible that LDs in this study have developed compensatory behaviors for their learning deficits, and utilize more visual information compared to auditory information [16]. In the most challenging listening conditions, the LDs may have attended primarily to the visual stimuli, in spite of the instructions to report what they heard. Support for this hypothesis can be found in recent work by Tiippana et al., showing that visual attention to the lips modulates perceptions of incongruent audiovisual speech [15].

The increased perception of visual stimuli by the LDs compared to the NLs suggests that the neural mechanisms underlying audiovisual integration differ between NLs and LDs. These differences may have a subcortical origin. This is supported by the finding that LDs with delayed timing of brainstem responses reported fewer combinations and fusions than NL children with normal brainstem timing. This suggests that for some LD children, a neurophysiologic deficit at the brainstem level may impact their perception of audiovisual speech. Brainstem deficits could directly impact the representation of auditory speech in the cerebellar cortex, before the signal reaches the neocortex. Mathiak et al. [10] proposed that the right cerebellar cortex represents the temporal structure of speech. A disruption in timing here would be relayed to the auditory cortex, conceivably affecting the process of phonetic representation.

Alternatively, the differences in NLs' and LDs' audiovisual perception could reflect differences in patterns of cortical activity elicited by audiovisual speech stimuli. For example, Shaywitz et al. demonstrated increased activation in Broca's area accompanied by decreased activation in Wernicke's area, angular gyrus and striate cortex in dyslexic children performing phonological tasks with visual stimuli [13]. Furthermore, Broca's area, and other speech-motor specific areas, have been shown to be activated when viewing speech movements without hearing speech, perhaps through the activation of the audiovisual mirror neuron system [6]. In a study by Breznitz [1], LDs exhibited greater cortical asynchrony between the representation of visual and auditory stimuli than NLs. Taken together, the present findings and the literature raise several possible cortical and subcortical mechanisms that may impact audiovisual perception.

The findings reported here indicate that distinctive patterns of perceiving audiovisual speech may underlie the reading and learning deficits of some children. In turn, this suggests that in order to facilitate a child's acquisition of reading skills, attention should be paid to a child's ability to utilize and integrate audiovisual speech. The link between audiovisual perception and reading is further supported by an audiovisual training study conducted by Kujala and colleagues [20] in which learning-disabled children who trained with nonlinguistic audiovisual stimuli improved their reading skills. Further perceptual and neurophysiologic studies of audiovisual speech in normal and learning disabled children may ultimately lead to the development of alternative strategies for reading instruction based on an individual's neurophysiologic profile.

Acknowledgements

This research was supported by NIH NIDCD DC01510. We would like to thank the children and their families who made this study possible.

References

[1] Z. Breznitz, Asynchrony of visual-orthographic and auditory-phonological word recognition processes: an underlying factor in dyslexia, Read. Writ. 15 (2002) 15–42.

[2] B. de Gelder, J. Vroomen, Impaired speech perception in poor readers: evidence from hearing and speech reading, Brain Lang. 64 (1998) 269–281.

[3] G.F. Eden, J.W. VanMeter, J.M. Rumsey, J.M. Maisog, R.P. Woods, T.A. Zeffiro, Abnormal processing of visual motion in dyslexia revealed by functional brain imaging, Nature 382 (1996) 66–69.

[4] J.J. Godfrey, A.K. Syrdal-Lasky, K.K. Millay, C.M. Knox, Performance of dyslexic children on speech perception tests, J. Exp. Child Psychol. 32 (1981) 401–424.

[5] C. King, C. Warrier, E. Hayes, N. Kraus, Deficits in auditory brainstem pathway encoding of speech sounds in children with learning problems, Neurosci. Lett. 319 (2002) 111–115.

[6] E. Kohler, C. Keysers, Hearing sounds, understanding actions: action representation in mirror neurons, Science 297 (2002) 846–848.

[7] N. Kraus, D.B. Koch, T.J. McGee, T.G. Nicol, J. Cunningham, Speech-sound discrimination in school-age children: psychophysical and neurophysiologic measures, J. Speech Lang. Hear. Res. 42 (1999) 1042–1060.

[8] N. Kraus, T.J. McGee, T.D. Carrell, S.G. Zecker, T.G. Nicol, D.B. Koch, Auditory neurophysiologic responses and discrimination deficits in children with learning problems, Science 273 (1996) 971–973.

[9] B. Lovegrove, Dyslexia and a transient/magnocellular pathway deficit: the current situation and future directions, Aust. J. Psychol. 48 (1996) 167–171.

[10] K. Mathiak, I. Hetrich, W. Grodd, H. Ackermann, Cerebellum and speech perception: a functional magnetic resonance imaging study, J. Cogn. Neurosci. 14 (2002) 902–912.

[11] H. McGurk, J. MacDonald, Hearing lips and seeing voices, Nature 264 (1976) 746–748.

[12] K.R. Pugh, W.E. Mencl, B.A. Shaywitz, S.E. Shaywitz, R.K. Fulbright, R.T. Constable, P. Skudlarski, K.E. Marchione, A.R. Jenner, J.M. Fletcher, A.M. Liberman, D.P. Shankweiler, L. Katz, C. Lacadie, J.C. Gore, The angular gyrus in developmental dyslexia: task-specific differences in functional connectivity within posterior cortex, Psychol. Sci. 11 (2000) 51–56.

[13] S.E. Shaywitz, B.A. Shaywitz, K.R. Pugh, R.K. Fulbright, R.T. Constable, W.E. Mencl, D.P. Shankweiler, A.M. Liberman, P. Skudlarski, J.M. Fletcher, L. Katz, K.E. Marchione, C. Lacadie, C. Gatenby, J.C. Gore, Functional disruption in the organization of the brain for reading in dyslexia, Proc. Natl. Acad. Sci. USA 95 (1998) 2636–2641.

[14] P. Tallal, S. Miller, R.H. Fitch, Neurobiological basis of speech: a case for the preeminence of temporal processing, Ann. N. Y. Acad. Sci. 682 (1993) 27–47.

[15] K. Tiippana, T. Andersen, M. Sams, Visual attention modulates audiovisual speech perception, Eur. J. Cogn. Psychol. (2003) in press.

[16] C. von Karolyi, E. Winner, W. Gray, G.F. Sherman, Dyslexia linked to talent: global visual-spatial ability, Brain Lang. 85 (2003) 427–431.

[17] G. Wilkinson, Wide Range Achievement Test-3 (WRAT-3), Jastak Associates, Wilmington, DE, 1993.

[18] R. Woodcock, M. Johnson, Woodcock-Johnson Psycho-Educational Battery: Tests of Cognitive Ability (W-J PEB), DLM Teaching Resources, Allen, TX, 1977.

[19] R. Woodcock, M. Johnson, Woodcock-Johnson Psycho-Educational Battery-Revised: Tests of Cognitive Ability, DLM Teaching Resources, Allen, TX, 1989.

[20] T. Kujala, K. Karma, R. Ceponiene, S. Belitz, P. Turkkila, M. Tervaniemi, R. Naatanen, Plastic neural changes and reading improvement caused by audiovisual training in reading-impaired children, Proc. Natl. Acad. Sci USA 98 (2001) 10509–10514.