The fragile nature of the speech-perception deficit in dyslexia: Natural vs. synthetic speech

Leo Blomert* and Holger Mitterer

Faculty of Psychology, Department of Cognitive Neuroscience, Universiteit Maastricht, P.O. Box 616 6200 MD Maastricht, The Netherlands

Accepted 28 July 2003

Abstract

A number of studies reported that developmental dyslexics are impaired in speech perception, especially for speech signals consisting of rapid auditory transitions. These studies mostly made use of a categorical-perception task with synthetic-speech samples. In this study, we show that deficits in the perception of synthetic speech do not generalise to the perception of more naturally sounding speech, even if the same experimental paradigm is used. This contrasts with the assumption that dyslexics are impaired in the perception of rapid auditory transitions.
© 2003 Elsevier Inc. All rights reserved.

1. Introduction

It is well established that the success in learning to read is related to the performance on phoneme-awareness tasks (Adams, 1990; Catts, 1989; Wagner & Torgesen, 1987), and at least a large number of dyslexics show deficits in phonological-awareness tasks. Different researchers have argued that this is due to less well-defined phonological representations in dyslexics (e.g., Brady, 1997; Snowling, 2000). However, it has been argued that poorly defined phonological categories are in turn caused by an impairment in auditory perception (e.g., Tallal, 1980). Poor auditory perception supposedly leads to poor perception of information-bearing elements in speech sounds, which, in turn, leads to poorly defined phonological categories. According to this claim, first defended by Tallal, the underlying deficit in dyslexia is to be found in poor auditory abilities.

Evidence for such problems in the perception of information-bearing elements in speech sounds stems from studies using a categorical-perception paradigm. While the general claim of categorical perception, that within-category discrimination is worse than between-category discrimination, is in decline (see, e.g., Gerrits, 2001; Massaro, 1998; Schouten, Gerrits, & van Hessen, 2003), it is generally accepted that speech sounds are decided upon—as opposed to perceived—categorically. That is, if listeners have to judge CV syllables stemming from a continuum of speech sounds and differing in the onset of the second formant, most stimuli will be identified consistently as either /da/ or /ba/ and only a small subset of stimuli will be identified inconsistently over different trials. This leads to a sigmoid identification curve with a rather high maximum slope, hence ‘categorical perception,’ or more precisely, ‘categorical decision.’

A number of studies report that dyslexics deviate from this pattern and show shallower identification functions than normal-reading controls (Godfrey, Sydral-Lasky, Millay, & Knox, 1981; Hurford & Sanders, 1990; Mody, Studdert-Kennedy, & Brady, 1997; Reed, 1989; Tallal, 1980). An obvious interpretation for this deficit is that dyslexics are impaired in perceiving the acoustic information bearing elements, especially transients (Stein & Walsh, 1997; Tallal, 1980; Tallal, Miller, & Fitch, 1993). A number of studies (Mody et al., 1997; Nittrouer, 1999; Rosen & Manganari, 2001), however, seriously questioned the assumption that the categorical-perception deficit associated with dyslexia can be attributed to ‘poor perception.’ Most notably, Serniclaes, Sprenger-Charolles, Carre, and Demonet (2001) report that dyslexics are better in discriminating within-category differences than normal-reading controls.

* Corresponding author.
E-mail address: l.blomert@psychology.unimaas.nl (L. Blomert).

This result renders the assumption that dyslexics are impaired in the perception of acoustic transients rather implausible. If dyslexics are impaired in perceiving acoustic transients, then their sensitivity to within- *and* between-category differences should be smaller than that of normal-reading controls.

Most studies investigating speech perception in dyslexics relied on synthetic speech (Hurford & Sanders, 1990; Reed, 1989; Tallal & Piercy, 1974, 1975). This is a defendable experimental choice, as synthetic speech is under complete control. The experimenter is free to vary just one parameter and assess the sensitivity of the listener to this particular parameter. However, it remains open to question whether results obtained with synthetic speech may be generalised to ecological speech perception. Iverson (2001), for instance, found that identification and discrimination performance of cochlear-implant users on a synthetic /da-/ta/ continuum only weakly predicted the comprehension of natural speech. Given this result, it is also questionable whether a categorical-perception deficit in dyslexics obtained with synthetic speech does generalise to the categorical perception of more naturally sounding speech.

Therefore, we compared the performance of dyslexic and control participants for natural and synthetic speech. To this end, we generated two continua based on natural speech and synthetic speech. We examined whether we could, first, replicate a categorical-perception deficit with stimuli comparable to those of Reed (1989) and, second, whether this deficit would generalise to a natural-speech continuum.

2. Method

2.1. Participants

Ten participants were recruited from the Regional Institute for Dyslexia (RID) in Arnhem and Maastricht, The Netherlands. These 10 participants had been diagnosed as dyslexic prior to the present study by the RID, which is one of the major specialised dyslexia institutes in The Netherlands. In addition 12 normal reading children were recruited from public schools.

These 12 control children were tested on two reading tests and an unpublished phoneme-deletion task (devised by the RID). Children in the control group were judged by their classroom teachers to be average or above average in academic achievement. In addition, all children had to pass a hearing test at 0.5. 1, 2, and 4 kHz at 20 dB and to perform within 15 dB of the norm level in a speech audiogram in order to be included in the study. The descriptive data are shown in Table 1.

The dyslexic participants had undergone an extensive cognitive diagnosis at the dyslexia institute. This testing procedure included an intelligence test (WISC-R, Dutch version; WISC-R projectgroep, 1986), two reading tests and an unpublished phoneme-deletion task (devised by the RID). The phoneme deletion test asks children to delete the first phoneme of an auditory presented word and pronounce the remaining word.

Dyslexic children were selected on the following criteria: Performance had to be at least one standard deviation below the age-appropriate mean on two standardised reading tests (the Een-minuut-leestest [one-minute-reading test], Brus and Voeten, 1972, and the KLEPEL, a pseudoword reading test devised by van den Bos, Spelbert, Scheepstra, & de Vries, 1994). Moreover, performance had to be within one standard deviation of the age-appropriate mean on the Wechsler Intelligence Test (WISC-R) and on a standardised visual-form perception test. In addition, the dyslexic children had to pass the same hearing tests as the control children in order to be included in the study. Performance on the word-reading and nonword-reading tests showed that the normal-reading-group performed at an age-appropriate level, while the dyslexic group performance lagged, respectively, two and one-and-a-half years behind. Performance on the phonological test showed that all dyslexics suffered from phonological problems.

2.2. Materials

Two speech-sound continua were generated. The first continuum ranged from /ta/ to /ka/ and was created

Table 1
Descriptive data for dyslexic and normal-reading group

	Group		$t(df = 20)$
	Dyslexic	Normal-reading	
Age (yr.mth): mean (range)	9.0 (8.0–9.10)	9.3 (8.8–9.8)	
WISC-R	107 (9.2)	–	
Reading level words[a]	5.1 (2.6)	9.9 (2.5)	4.4*
Reading level nonwords[a]	6.1 (2.4)	10.2 (2.8)	3.6*
Phoneme deletion[b]	18.8 (6.1)	25.6 (2.2)	3.6*

* ($p < .01$).
[a] Standardised scale; mean = 11, $SD = 2$.
[b] Number of correct items (max = 28).

from a natural utterance of /ta/ by a female speaker of standard Dutch. Formant estimations of the natural utterance were made using the software package PRAAT 3.9 (Boersma & Weenink, 1999), and the stable phase of the vowel was found to start 80 ms post-onset. For these 80 ms, the third formant was manipulated by lowering the onset frequency from the original 3.3–2.4 kHz in six steps of 0.33 each. In order to minimise residual cues to place of articulation in the noise source, a low-pass filtered synthetic noise source was used. The edited stimuli retained the naturalness of the original utterance.

The other continuum was based on the synthetic source-filter model provided by the PRAAT software package. It ranged perceptually from /ba/ to /da/. The source had a length of 0.2 s and started at 120 Hz, rose to 130 Hz at 40 ms and then fell to 110 Hz at 0.2 s. The filter consisted of five formants with a steady state (and bandwidth, all values in Hz) at 750 (100), 1200 (150), 2350 (200), 3300 (300), and 4000 (500). The upper two formants were constant in all stimuli. The lower three formants for the /ba/ stimulus had onset frequencies of 300, 1100, and 2150 Hz, and reached the steady state at 0.05 s. The /da/ stimulus had the same first formant trajectory, while the second and third formant started at 1600 and 2600 Hz, respectively. Five intermediate stimuli were created by interpolating between these values using a bark scale. Using the bark scale instead of the Hz-scale leads to perceptually equivalent intervals on a continuum. This continuum sounded quite unlike natural speech due to the source characteristics and the stable formant parameters, which lack natural fluctuations.

2.3. Design and procedure

The data for off-line screening and the experimental procedures were collected over the course of two sessions on different days. Before the experimental procedure, there was a cognitive screening including the reading tasks and the hearing tests. Then, participants were acquainted with the two alternative forced choice (2AFC) procedure using the salient Dutch minimal pair 'huis'-'muis' [house-mouse]. Two pictures representing the two words were displayed on the computer screen on the upper left and upper right corner. Participants were instructed that, upon hearing a word, they should indicate which word they heard by pressing an upper left ('q') or an upper right ('p') key of the computer's keyboard. (A custom-made cover left only these two keys available for pressing.) After each reaction, a 'smiley' face indicated whether the choice was correct. None of the children displayed any problems in understanding the 2AFC task. After this, children performed a 2AFC task with the two endpoints of the semi-synthetic [tart]-[kart] continuum. Two pictures on the computer screen indicated which key was associated with which word. If more than four errors were made on the first 20 trials, the training session was repeated. Then the main experimental session started. In this task, all seven semi-synthetic stimuli were presented to be categorised without feedback. Every stimulus was presented ten times, order was randomised. This first experimental task lasted about 15 min.

In the second session, the same procedure of training and experimental session was repeated now using the synthetic /ba/-/da/ stimuli. First, there was a training phase using the two endpoints of the continuum in which explicit feedback was provided. Then all stimuli, including the intermediate stimuli were judged ten times each without feedback. The order of presentation was randomised.

3. Results

The results were coded in percentage /ta/ and /ba/ responses in the synthetic and natural stimulus condition, respectively. These measures were subjected to a repeated-measure ANOVA with Stimulus Type (natural vs. synthetic) and Continuum (the seven steps of each continuum) as within-subject variables and group (dyslexics vs. normal-reading) as between-subjects variable (see Fig. 1). The analysis revealed a significant main effect of Continuum ($F(6, 120) = 139.9, p < .001$), while the other main effects did not reach significance ($Fs < 1$). All two-way interactions were significant (Stimulus

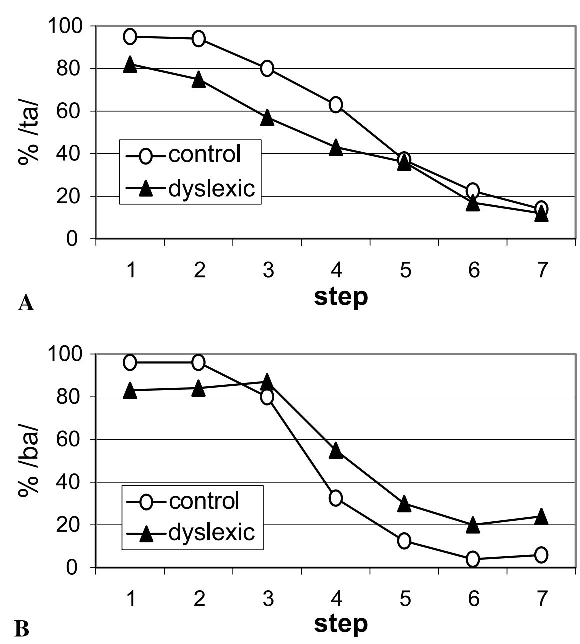

Fig. 1. Identification functions of dyslexic and normal-reading groups on (A) a speech-sound continuum based on natural speech and (B) a speech-sound continuum based on synthetic source-filter modelling.

Type-by-Group: $F(1,20) = 6.4, p < .025$, Continuum-by-Stimulus Type: $F(6,120) = 4.7, p < .001$, Continuum-by-Group: $F(6,120) = 3.3, p < .01$). However, the two-way interactions were qualified by a significant three-way interaction ($F(6,120) = 2.2, p < .05$). In order to investigate the nature of the three-way interaction, separate ANOVAs were calculated for both Stimulus Types using Group and Continuum as predictors.

Using the data from the natural-speech continuum, there was a significant effect of Continuum ($F(6,120) = 66.3, p < .01$), but neither a significant effect of Group $F(1,20) = 3.1, p = .09$ nor a significant interaction between Group and Continuum ($F(6,120) = 1.6, p > .15$). Then the data from the synthetic-speech continuum were assessed using a similar ANOVA. This also revealed a significant effect of Continuum ($F(6,126) = 71.8, p < .01$). However, both the main effect of Group ($F(1,21) = 6.0, p < .025$) and its interaction with Continuum ($F(6,120) = 4.2$, $p < .0025$) were significant. To further investigate the interaction, the effect of Group was tested at each level of the continuum. It turned out that the normal-reading group perceived the first two steps of the continuum more consistently than the dyslexic group as /b/ ($t(20) = 2.5, p < .025; t(20) = 2.3$, $p < .05$), while perceiving more consistently /d/ at the last three steps of the continuum ($t(20) = -2.6, p < .025$; $t(20) = -2.4, p < .05$; $t(20) = -2.8, p < .025$). At steps three and four, there was no significant group effect ($t(20) = -1.1, p > .2$; $t(20) = -1.8, p = .091$).

Individual identification performance is often assessed by estimating the maximal slope of the identification functions. To this end, we estimated the maximal slope of the individual identification functions by fitting a logistic function with two free parameters, a slope and a bias parameter to the individual identification functions. The slope parameters (see Table 2) were used in a repeated-measure ANOVA with Stimulus Type (natural vs. synthetic) as within-subject variable and group as between subject variable. This analysis revealed a significant main effect of Stimulus Type ($F(1,20) = 6.3, p < .025$), but no main effect of Group and no interaction ($Fs < 1$). The main effect of Stimulus Type shows that the synthetic continuum is perceived more categorically than the natural continuum.

Table 2
Mean (standard deviation) estimated slopes of individual identification functions for the natural and the synthetic continuum in the dyslexic and normal-reading group

Group	Continuum	
	Natural	Synthetic
Normal-reading	4.42 (2.46)	9.28 (9.28)
Dyslexic	2.81 (1.42)	8.15 (8.15)

4. Discussion

The current study aimed to examine in how far the assumed speech-perception deficit in dyslexia generalises to more naturally sounding speech. We used a synthetic-speech continuum of the same type as typically used in similar studies over the last two decades. With these stimuli, we replicated earlier results with similar stimuli that indicated a speech-perception deficit in dyslexia. This deficit was not evident in the slopes of the identification curves but rather in less consistent responses at the endpoints of the continuum in the dyslexic group. However, no significant differences were found when we used the same paradigm but a stimulus continuum based on natural speech. No speech-perception deficit in dyslexia was observed with these stimuli.

One problem with interaction findings generally is that they may be due to a ceiling effect. That is, a difference found in one condition is not found in another condition, because performance is at ceiling in that other condition. Note that such an explanation cannot apply to the present data. This is due to the fact that a categorical-perception deficit was only found in the easier to categorise continuum, the synthetic speech continuum.

That the natural-speech continuum was more difficult to categorise for both groups is evident from the analysis of the slope parameters. This raises the question why the natural-speech continuum is more difficult to categorise, as evident from the smaller slope parameters. Naively, one may assume that natural speech is easier to understand than synthetic speech. That is normally the case, because natural speech contains multiple and redundant cues for each feature distinction. Because of this redundancy, the speech signal remains understandable, even if it is highly distorted as in sine-wave speech, noise-vocoding, or reduced to transient signals (Remez, Rubin, Pisoni, & Carrell, 1981; Scott, Blank, Rosen, & Wise, 2000; Shannon, Zeng, Kamath, Wygonski, & Ekelid, 1995). However, not all of these cues can be captured in LPC-based formant estimations, and, therefore, residual cues to the original place of articulation will remain in the estimated source. These residual cues indicating the original place of articulation will conflict with the manipulated formant transition and, as a result, the natural-speech continuum is more difficult to perceive. In contrast, the synthetic source is completely neutral in terms of place cues. This leads to a well-controlled, one-cue-varying synthetic-speech continuum, which is easier to categorise.

It is also noteworthy that the moderation of the speech-perception deficit by stimulus type is opposite to the direction that might be expected on the basis of a poor-auditory-perception hypothesis. If this would be the case, the more difficult-to-perceive continuum—in this case, the natural-speech continuum—should lead to stronger group differences. Opposite to that prediction, the more natural sounding stimuli produced less pronounced group differences.

This analysis of the difference between synthetic and natural speech also suggests an explanation for the data indicating a speech-perception deficit in dyslexia (Godfrey et al., 1981; Hurford & Sanders, 1990; Reed, 1989; Tallal, 1980). Instead of being poor in perceiving speech stimuli, dyslexics might be less able to adept to the range of novel stimuli they hear in a categorical-perception task with synthetic stimuli (see, e.g., Ladefoged and Broadbent, 1957, for evidence for such a process of speaker normalisation). While both groups seem not to differ in the 'analogue neural representation' of the formant transition, the control group is better able to apply their phonological categories built on natural speech consistently to the novel synthetic stimuli (for similar claims, see Brady, 1997; Snowling, 2000; Studdert-Kennedy, 2002).

Some have argued that the poor performance of language-impaired children (SLI) on perceptual tasks is due to meta-cognitive task requirements as sustained attention and memory (Bishop, Carlyon, Deeks, & Bishop, 1999). Note that the present data show that a similar interpretation of the performance of dyslexic children on a categorical-perception task does not apply. If task requirements would play a role, then dyslexics should be impaired in all categorical-perception tasks. However, a slight difference in stimulus quality was enough in the present study to obtain a contrast between normal and impaired behaviour in dyslexic participants on a categorical-perception task. This is difficult to explain in terms of task requirements.

It is important to note that the naturalness of the continua in the present study may be confounded with the type of contrast tested (labial to alveolar versus alveolar to velar). Nevertheless, both contrasts were place-of-articulation contrasts and hence relied on the manipulation of formant transitions. Therefore we should have observed an impairment for both contrasts if dyslexics were impaired in the perception of formant transitions.

In summary, our results contribute to a growing body of data that contradict the assumption that dyslexics have a deficit in the perception of short, acoustic transients (Mody et al., 1997; Nittrouer, 1999; Rosen & Manganari, 2001; Serniclaes et al., 2001). If such a deficit were present, a speech-perception deficit should be observed with synthetic as well as natural speech continua based on manipulating formant transitions. Instead, the present results are compatible with a 'phonological-coding deficit.' According to such an account, "the deficit is in the phonetic transform from analogue neural response patterns to digital lexical/phonological representations" (Studdert-Kennedy, 2002, p. 6)

Acknowledgments

The authors wish to thank the Regionaal Instituut voor Dyslexie (RID) and the Basisschool Wyck, Maastricht for their support in acquiring participants. The RID also provided us with the diagnostic data of the dyslexics. In addition, we like to thank Christiaan Paffen for his assistance in running the experiments. This research was partly supported by Grant 048.011.046 of the Netherlands Organization for Scientific Research to the first author.

References

Adams, M. J. (1990). *Beginning to read: Thinking and learning and about print*. Cambridge, Mass: MIT Press.

Boersma, P., & Weenink, D. (1999). *PRAAT 3.9: a system for doing phonetics with the computer [Computer software]*. Amsterdam, The Netherlands: University of Amsterdam.

Brady, S. (1997). The ability to encode phonological representations: An underlying difficulty of poor readers. In B. A. Blachman et al. (Eds.), *Foundations of Reading Acquisition and Dyslexia: Implications for Early Intervention* (pp. 21–47). Mahwah, NJ: Lawrence Erlbaum Associates.

Bishop, D. V. M., Carlyon, R. P., Deeks, J. M., & Bishop, S. J. (1999). Auditory temporal processing impairment: Neither necessary nor sufficient for causing language impairment in children. *Journal of Speech, Language, and Hearing Research, 42*, 1295–1310.

Catts, H. W. (1989). Defining dyslexia as a developmental language disorder. *Annals of Dyslexia, 39*, 50–65.

Gerrits, P.A.M., 2001. The categorization of speech sounds by childrens and adults: a study of the categorical perception hypothesis and the developmental weighting of acoustic speech cues [dissertation]. Utrecht: LOT.

Godfrey, J. J., Sydral-Lasky, A. K., Millay, K. K., & Knox, C. M. (1981). Performance of dyslectic children on speech perception tests. *Journal of Experimental Child Psychology, 32*, 401–424.

Hurford, D. P., & Sanders, R. E. (1990). Assessment and remeditation of a phonemic discrimination deficit in reading disabled second and fourth graders. *Journal of Experimental Child Psychology, 50*, 396–415.

Iverson, P., 2001. Individual differences in phonetic perception by adult cochlear implant users: Effects of sensitivity to /d/-/t/ on word recognition. *Speech, Hearing, and Language: work in progress, 13*, 1–21. London: UCL.

Massaro, D. (1998). *Perceiving talking faces: from speech perception to a behavioral principle*. Cambridge, MA [etc.]: MIT Press.

Mody, M., Studdert-Kennedy, M., & Brady, S. (1997). Speech perception deficits in poor readers: Auditory processing or phonological coding? *Journal of Experimental Child Psychology, 58*, 199–231.

Nittrouer, S. (1999). Do temporal processing deficits cause phonological problems? *Journal of Speech, Language, and Hearing Research, 42*, 925–942.

Reed, M. A. (1989). Speech perception and the discrimination of brief auditory cues in reading disabled children. *Journal of Experimental Child Psychology, 48*, 270–292.

Remez, R. E., Rubin, D. B., Pisoni, D. B., & Carrell, T. D. (1981). Speech perception without traditional cues. *Science, 212*, 947–950.

Rosen, S., & Manganari, E. (2001). Is there a relationship between speech and nonspeech auditory processing in children with dyslexia? *Journal of Speech, Language, and Hearing Research, 44*, 384–388.

Schouten, B., Gerrits, E., & van Hessen, A. (2003). The end of categorical perception as we know it. *Speech Communication, 41*, 71–80.

Scott, S. K., Blank, C. C., Rosen, S., & Wise, R. J. (2000). Identification of a pathway for intelligible speech in the left temporal lobe. *Brain, 123*, 2400–2406.

Serniclaes, W., Sprenger-Charolles, L., Carre, R., & Demonet, J.-F. (2001). Perceptual Discrimination of Speech Sounds in Developmental Dyslexia. *Journal of Speech, Language, and Hearing Research, 44*, 384–388.

Shannon, R. V., Zeng, F. G., Kamath, V., Wygonski, I., & Ekelid, M. (1995). Speech perception with primarily temporal cues. *Science, 270*, 303–304.

Snowling, M. J. (2000). *Dyslexia* (second ed.). Oxford, UK: Blackwell Publishers.

Stein, J., & Walsh, V. (1997). To see but not to read: The magnocellular theory of dyslexia. *Trends in Neuroscience, 20*, 147–152.

Studdert-Kennedy, M. (2002). deficits in phoneme awareness do not arise from failures in rapid auditory processing. *Reading and Writing, 15*, 5–14.

Tallal, P., & Piercy, M. (1974). Developmental aphasia: Rate of auditory processing and selective impairment of consonant perception. *Neuropsychologia, 12*, 83–93.

Tallal, P., & Piercy, M. (1975). Developmental aphasia: The perception of brief vowels and extended stop consonants. *Neuropsychologia, 12*, 69–74.

Tallal, P. (1980). Auditory temporal perception, phonics, and the reading disabilities in children. *Brain and Language, 9*, 182–198.

Tallal, P., Miller, S., & Fitch, R. H. (1993). Neurobiological basis of speech: a case for the preeminence of temporal processing. *Annals of the New York Academy of Sciences, 682*, 27–47.

Wagner, R. K., & Torgesen, J. K. (1987). The nature of phonological processing and its causal role in learning to read. *Psychological Bulletin, 101*, 192–212.

WISC-R Projectgroep: Haasen v. P.P., De Bruyn E.E.J., Pijl Y.J., Poortinga Y.H., Spelberg H.C., Van der Steene G., Coetsier P., Spoelders-Claes & Stinissen J., 1986. *WISC-R. Nederlandstalige uitgave*, Lisse: Swets & Zeitlinger.