AUDIOVISUAL SPEECH PERCEPTION IN DYSLEXICS: IMPAIRED UNIMODAL PERCEPTION BUT NO AUDIOVISUAL INTEGRATION DEFICIT

R Campbell, A Whittingham, U Frith with *DW Massaro, MM Cohen*
University College London, University of California at Santa Cruz

R. Campbell, Department of Human Communication Science
University College London WC1 N 1 PG
e-mail R.Campbell@ucl.ac.uk phone +44 (0) 171 837-0113

ABSTRACT

Dyslexic and control school students (mean age 13.5 years) were tested using a series of unimodal and bimodal speechtokens; ba,va,tha,da and ga in every combination and unimodally. They made significantly more errors than controls to unimodal auditory and unimodal visual tokens. Their integration of audiovisual speech was not significantly impaired, but it did show a different function than that of controls, reflecting the anomalous unimodal processing functions. These data fit recent suggestions that magnocellular function is impaired in dyslexia. They fit less easily both with auditory temporal processing deficit theories or with theories of defective phonemic processing . A general 'integration-deficit' theory (Paulesu *et al*, 1996) is not supported by these data.

1. INTRODUCTION

Dyslexic people fail at reading and spelling despite normal intelligence, good schooling and effective home support. Developmental dyslexia is increasingly recognized as an organic condition: it runs in families and it is associated with distinctive structural brain morphology and with specific anomalies of processing under cognitive and neurophysiological test. In general, dyslexic youngsters show some difficulties in acquiring spoken language. Thus, many theoretical approaches stress the failure of the dyslexic student to develop reliable phonological representations of speech: representations needed to understand the alphabetic principle (the mappings of speech sounds to their letters). One approach (Tallal, 1980) suggests that the primary deficit is in acoustic perception: the dyslexic cannot resolve fast changing auditory temporal events which might help distinguish acoustic formants. Another (Liberman & Liberman, 1990) contends that acoustic deficits are irrelevant: phonological processing is amodal and *sui generis*. A brain imaging study (Paulesu *et al*, 1996) has shown a structural correlate of the often reported dyslexic problems with (spoken) rhyme and immediate memory tasks .While normal readers showed activity in the **insula** for these speech tasks, this was absent in dyslexics. The insula is a cortical region intermediate between frontal and temporal lobes: Paulesu *et al*. suggest that it is critical to integration of letter and speech-sound knowledge. But a number of other suggestions have been made concerning neurobiological contributions to dyslexia, including the suggestion that there is a deficit in magnocellular visual function in dyslexia. The magnocellular system shows fine temporal resolution but relatively poor spatial resolution, and poor discrimination of hue. The contribution of this system to reading may lie in the control of eye movements: both saccades and vergence (Stein & Walsh, 1997).
The performance of dyslexics on tasks of visual (speechreading) and of audiovisual syllable identification, including fusion effects (McGurk & MacDonald, 1976) is of interest in this context. Auditory discrimination theory predicts no difficulty in silent speechreading. The theory of deficient integration suggests that McGurk effects should be reduced in dyslexics and that congruent audiovisual inputs need not be processed more accurately than unimodal inputs. The magnocellular theory might predict difficulties in purely visual speechreading. Neurioimaging work indicates that area V5, one of the major cortical projection areas for the transient (magnocellular)system, together with neighbouring superior temporal cortex, is implicated in silent speechreading and the perception of biological motion (Calvert *et al*, 1997; Howard, 1996).
A previous study by de Gelder & Vroomen (see Vroomen, 1992; de Gelder & Vroomen, 1995) explored this topic using a natural face image seen to be saying 'da' or 'ba', synchronized to an auditory token taken from a nine-step synthesized speech token ranging from 'ba' to 'da' (Massaro & Cohen, 1983). Vroomen found dyslexic children were poorer than control groups at silent speechreading (ba or da classification) and were more variable in their classification of the ninestep auditory continuum. As for audiovisual integration '*..no significant difference of visual influence, although there is trend in the direction that dyslexics are less influenced by vision than control subjects..' (p 145).*

We used a different stimulus display (see below) and older students to attempt a more detailed investigation of this issue.

1.1 Method

A videotape comprising auditory, visual and audiovisual tokens of the monosyllables /ba/,/va/,/tha /da/, /ga/ was constructed using a natural male speaker and a synthetic face, and each face appeared unpredictably within each trial run. The auditory and the natural video tokens were taken from a single male speaker on the Bernstein & Eberhardt videodisk.
For synthetic visible speech, a parametrically controlled polygon topology was used to generate the syllables. The facial surface was modelled as a polyhedral object joined at the edges with a smoothshaded surface. Further details of the synthetic face and of the dubbing procedure for fitting the speaking face to the synthesized auditory speech tokens, are in Campbell *et al* (1997).

A complete trial set comprized every experimental combination of 50 dubbed syllables (5 natural faces, 5 synthetic, all dubbed to the 5 speechsounds) and 10 unimodal (5 vision alone, 5 audition alone) tokens, all presented in random order. Each subject saw six such trial sets. Within each trial set the order of tokens was randomized. The experiment lasted forty-five minutes for each subject
Instructions to subjects were to watch and listen to the speaker and to report the syllable 'that you think you heard': a full response choice was indicated; that is, subjects were told that the syllable could be any of 'ba, va, tha, da, ga - or a combination or blend of these - such as a consonant cluster-vowel combination'. The image was shown on a large colour TV monitor and subjects were seated about 1 metre from the screen. For unimodal auditory tokens no image was seen on the dark screen Loudness level was set at 65db and testing was in a quiet room. Subjects were tested in ones or in twos, and their responses were written. Subjects were mainstream students, all volunteers from a single London secondary school. All the dyslexic group had been assessed as dyslexic by the educational psychologist.

1.2 Subject characteristics

means (sds)	controls (n=8)	dyslexics (n=10)	t-test p
age in months	169 (18.04)	169 (11.19)	ns
reading age	163.3 (5.30)	128.7 (21.08)	0.01
spelling age	153.75 (20.0)	128.50 (9.89)	0.001
digit span	11.62 (1.41)	10.20 (1.31)	0.05
Block design WISC	12.25 (2.12)	13.4 (2.22)	ns
Vocab WISC	9.75 (2.61)	11.10 (1.63)	ns
spoonerisms (time in sec)	55.37 (6.65)	133.0 (55.22)	0.001

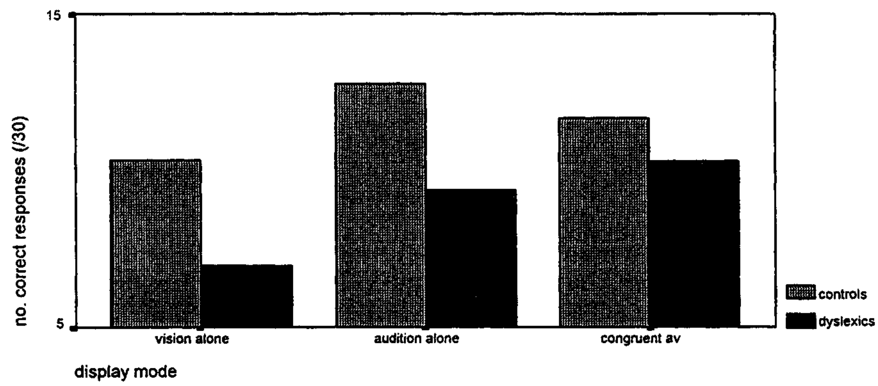
Subject characteristics are summarized in the table above. These show mean standardized reading and spelling scores, performance on digit span test, two items of the standardized IQ test (WISC) and on a timed test of spoonerisms (exchanging the initial sound of two words thus: 'tore foe' > 'four toe')

2. RESULTS

2.1 Correct responses

For unimodal and for congruent bimodal conditions, the mean number of completely correct responses was calculated. These are summarized in figure 1.

figure 1 mean correct responses unimodal and congruent conditions



ANOVA with within-subject factors of modality (vision alone, audition alone or audiovisual-congruent) and type of utterance (ba,va,tha,da,ga), and dyslexia as the between-subjects factor was performed. All three main factors were significant. The effect of group (see above) was significant at $p < 0.01$. It failed to interact with modality. That is, dyslexics were poorer than controls whether silently seen, heard or congruent av tokens were shown. The main effect of mode reflects that vision alone was worse than either audition or congruent av. The interaction of type and modality (not shown here) was highly significant, and failed to interact with group. While speechreading was hardest for 'tha' and for 'ga', this was not so for hearing.

2.2 Errors

Error-type was analysed for the three modalities of presentation. For vision-alone, the main effect of group was moderated by an interaction: dyslexics made significantly more omissions and 'other' responses (that is responses not indicated in the instructions) than controls. A similar pattern characterized audiovisual errors. Audition alone did not give a distinctive error pattern for each group.

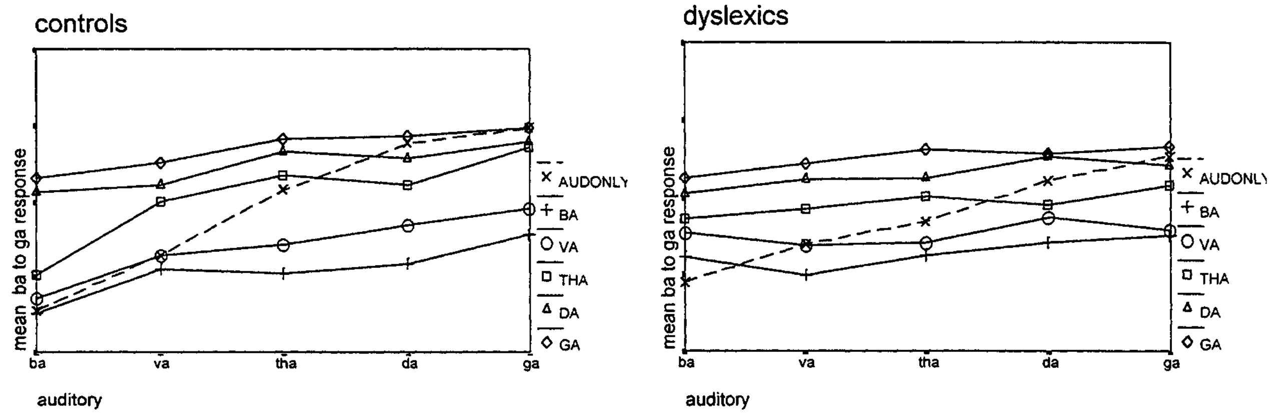
For incongruent AV tokens there is no 'correct' or 'incorrect' response. However since the series 'ba' to 'ga' is an ordinal series with respect to place of articulation, the series can be numerically coded to reflect this.

Using values from 1 to 5 to code ba, va, tha, da and ga respectively (with codes of 3 for blends such as 'dba', and coding for place of articulation for other responses, so that 'ta' was coded identically to 'da'), analyses of the effect of vision on audition can be performed.The relevant functions are shown with respect to audition alone in the figures below. The horizontal axis is auditory token identity; the vertical axis the mean coded value of the response as a number between 1 (ba) and 5 (ga).

2.3 Incongruent audiovisual speech, and baseline performance - coded responses

Figure 2 Coded audiovisual responses

Effects of each visual token on each auditory token



2.31 Analyses

A general ANOVA with two between factors (auditory type and visual type) each of which had five levels, and group, gave the following significant results. There were main effects of visual type (F (64,4)=17.23, p< 0.001), of auditory type (F(64,4)= 94.83, p < 0.001). There was also a significant interaction of group with visual type (F (64,4)= 4.85, p < 0.01). No other main effects or interactions approached significance. Separate ANOVAs for each group probed these differences further. For controls, the main effects of visual type, of auditory type, and their interaction, were significant at p< 0.01 or above (F (28,4)=22.30; F (28,4)=87.85 and F (112,16) =2.91, respectively). For dyslexics, only the effect of auditory type was significant (F(36,4)=30.56, p < 0.001. Neither the effect of visual type (F(36,2)=2.01, p= 0.08) nor the interaction (F(144,16)= 0.67) were significant.

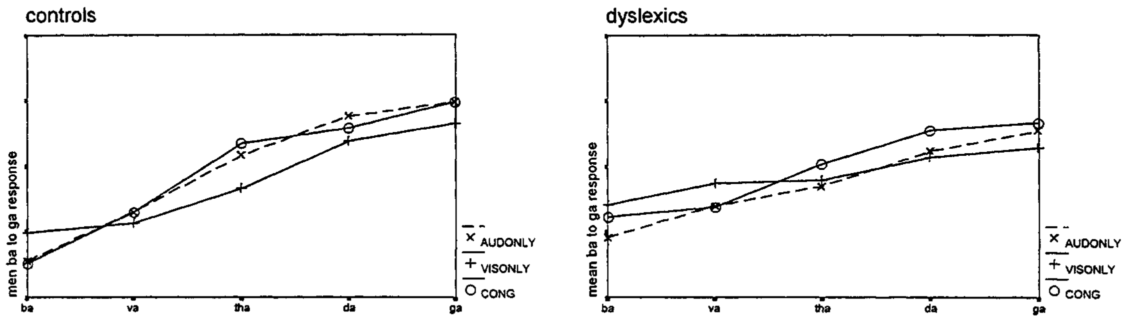
While both groups are affected by vision in their responses, the dyslexic group shows a reduced level of constraint from vision on report of auditory tokens. (flatter functions in the figure above). However, it cannot be concluded that this indicates a lack of integration of auditory and visual tokens. The vision-alone function for dyslexics is also flatter (less accurate) than that for controls. Thus, the gain in accuracy of report from adding vision to audition in this group may not be very marked.

2.4 Congruent and unimodal conditions - analysis of coded responses

Figure 1 shows dyslexics can benefit from bimodal inputs: does this occur for coded responses, too? Figure 3 shows the relevant coded functions.

ANOVA (factors - group, mode (3 levels) and type (5 levels)) shows no main effect of group, significant effects of mode (F (96,6) = 12.23, p< 0.001) and of type (F(64,4)= 100.58,p< 0.001) and interactions of each of these factors with group (F mode x group (96,6) = 12.23, p< 0.01; F type x group (64,4) = 3.69, p< 0.01), in addition to the mode x type interaction (F (384,24)= 1.70 , p< 0.05). However, there were no further interactions with group.

Figure 3. Coded unimodal and av-congruent mean responses



3. DISCUSSION

Confirming and extending de Gelder & Vroomen's (1995) findings, dyslexics are worse than controls at identifying auditory and visual syllable tokens. Because of this, their audiovisual performance is somewhat different to that of controls. Dyslexics nevertheless benefit from congruent audiovisual inputs, and we find no statistical evidence to suggest that their integration mechanisms are defective. This is not a trivial conclusion. Some individuals with impaired cortical vision **cannot** integrate seen and heard speech and their bimodal performance can be **poorer** than that in either modality alone (de Gelder *et al*, 1997).

The implications of this are practical and theoretical. Practically, they suggest that the delivery of natural speech bimodally could be especially useful for dyslexics. Natural visual and auditory speech deliver complementary phonetic information, so speechreading may make good any lack in auditory perception. Theoretically, these data do not support the claim that a general integration deficit underlies dyslexia. Integration mechanisms that fail in dyslexia may be at a different level of representation than those that support seen and heard speech.

Since both speechread and heard token identification is weak in dyslexics, a 'purely auditory' theory is inadequate. We might predict that if pure auditory discrimination difficulties underlie dyslexia, young dyslexics would try to make more use of natural lipreading and should have become proficient in using this supporting modality in developing their representations of speech. That this does not happen suggests the deficit is at a different level of function. The theory that phonological representations are deficient in dyslexics might predict that hearing, speechreading **and** av speech should all be similarly affected. Yet congruent av speech improves with respect to the unimodal condition, suggesting a true **perceptual** difficulty, but one not limited to the auditory modality. One likely candidate for perceptual frailty is the magnocellular visual system, whose major cortical projections play an important role in silent speechreading (Calvert et al, 1997), and which has been implicated in other cross-sensory systems (Livingstone et al., 1991).

4. REFERENCES

Calvert G. *et al* 1997 Activation of auditory cortex during silent speechreading *Science* **276** 5312,593-596

Campbell R. 1996 Seeing Brains Seeing Speech In D.Stork & M.Henneke (Eds) S*peechreading by machines and humans* Springer, Berlin, 115-135

Campbell R. *et al* 1997 Speechreading in the akinetopsic patient LM *Brain* **120,** 000-000

De Gelder B. *et al* 1997 Impaired Speechreading and Audio-visual Speech Integration in prosopagnosia. In Campbell R, Dodd BJ & Burnham D (Eds) *Hearing by Eye II* Psychology Press, Hove, 000-000

de Gelder,B & Vroomen J 1995 Memory deficits for heard and lipread speech in young and adult dyslexics In B. de Gelder & J.Morais (Eds*) Speech and Reading: a comparative approach* Lawrence Erlbaum125-139

Howard RJ. *et al* 1996 A direct demonstration of functional specialization within motion-related visual and auditory cortex of the human brain *Current Biology*_**6,** 1015-1019

Liberman IY. & Liberman A. 1990 Whole language versus code emphasis: underlying assumptions and their implications for reading instruction *Annals of Dyslexia,* **40** 51-76

Livingstone M. *et al* 1991 Anatomical and physiological evidence for magnocellular deficit in developmental dyslexia *Proc Nat Acad Sci,* **88** 7943-47

Massaro D. & Cohen MM. 1983 Evaluation and integration of visual and auditory information in speech perception *JEP HP&P,* **9** 753-771

McGurk H. & MacDonald J. 1976. Hearing lips and seeing voices. *Nature* **264** 746-748

Paulesu E. *et al* 1996. Is developmental dyslexia a disconnexion syndrome? *Brain* **119**, 143-157

Tallal P. 1980 Auditory temporal perception, phonics and reading disabilities in children *Brain & Language* **9** 182-198

Stein J. & Walsh V. 1997 To see but not to read; the magnocellular theory of dyslexia *TINS* **20**,147-152

Vroomen J. 1992 *Hearing Lips and Seeing Voices.* PhD thesis, University of Tilburg