AUDIOVISUAL SPEECH PERCEPTION IN WILLIAMS SYNDROME

M.Böhning [1], R. Campbell [2] and A.Karmiloff-Smith [3]

(1) Department of Linguistics, Unversity of Potsdam, Germany

(2) Department Of Human Communication Science, University College London

(3) Neurocognitive Development Unit, Institute Of Child Health, University College London

ABSTRACT

In the rare genetic disorder of Williams syndrome (WS), visuospatial abilities, including face processing can be impaired. Auditory speech processing, on the other hand, may be less so. Claims have also been made that WS may impair integrative processing. In this context, the exploration of visual and audiovisual speech perception in WS is of interest.

Tokens from a single natural English speaker of the form /∧ba:/, /∧va:/, /∧θa:/, /∧da:/ and /∧ga:/, were digitally manipulated and presented in unimodal (vision alone, audition alone) and audiovisual conditions, for participants to identify each token.

Compared with age-matched controls, WS participants were impaired at **visual** but not auditory identification, and in audiovisual testing showed correspondingly reduced effects of vision on report of auditory token identity. Audiovisual integration was nevertheless demonstrable in WS. Visual phoneme identification may require visual skills that do not reach age-appropriate levels in WS, despite their age-appropriate (auditory) phonological abilities.

1 INTRODUCTION

Williams syndrome (WS) is a genetic disorder with an incidence of one in 20,000 live births [18]. It is of interest to cognitive neuroscientists because of its uneven cognitive-linguistic profile. Some language and communicative skills appear relatively proficient, while many nonverbal skills, especially visuo-spatial cognition, number, as well as planning, conceptual/semantic skills and problem solving, are severely compromised [13,17]. Face processing is not spared. Although people with WS are sociable and show appropriate responses to a range of facial acts that denote emotion and intention [3,14], Deruelle and colleagues [7] found a developmental delay in face processing in people with WS. They showed a processing deficit for faces in line with performance by much younger children. This was confined to configural processing; that is to face perception that relied on configural relations between face parts. The ability to recognize face parts in isolation or when mono-oriented (i.e., inverted faces) was relatively spared. Deruelle and colleagues also tested ability to match still face images (which varied in identity) for lipshape (/a/, /o/ and /i/). This was the only task where WS achieved age-appropriate performance. It was concluded that lip-reading relies on local rather than global or configural processing. This would account for its sparing in WS.

This conclusion may be precipitate. Speechreading is sensitive to the normal facial configuration. Audio-visual illusions are affected by the orientation of the face, being reduced for inverted faces, even when the mouth region itself is upright [8,16]. Moreover, unlike the static stimuli used by Deruelle et al., the processing of faces for speech makes use of visual movement. Time-varying information in the *absence* of a recognizable image of the relevant face parts can affect auditory perception [15].

The first aim of the study reported here was to describe the pattern of influence of seen on heard speech identification in people with WS, which has not yet been reported. Additionally, an hypothesis may be tested. A high-level integration deficit, which would impair the analysis of the products of perceptual processing, has been proposed as one cause of the anomalous cognitive profile in WS [e.g. 17]. The extent to which cross-modal perceptual integration might be implicated has not been explored hitherto, and audiovisual speech processing offers a suitable testbed. It is established in infancy, is largely insensitive to attentional and strategic demands, and follows a well-defined integration metric. Its cortical basis is also becoming clear [10-12,4].

Our experimental questions were: (1) How good is visual speech discrimination for a subset of syllables in WS? Here predictions were open. (2) How good is auditory speech discrimination? We had no reason to believe this to be compromised. (3) Finally, do WS show defective integration, with reduced effects of vision on audition?

2 METHOD

Participants were thirteen people with WS (diagnosed genetically by FISH technique), pre-screened for suitability for testing. Five were male. Age range was 11;1 - 52;2 years. A range of psychometric data, including full IQ measures were obtained for this group. Controls were matched individually for age and gender to the WS participants, and were recruited through a youth club and from volunteers in the London area.

Stimuli were derived from five naturally spoken iambic VCV syllables (/∧ba:/, /∧va:/, /∧θa:/, /∧da:/, and /∧ga:/) spoken by a female British-English speaker. These were digitally captured, matched for overall length and for perceived onset of the consonant, then spliced into audio- and visual-only segments. These were used to generate 35 tokens: 5 auditory, 5 visual and the complete set of 25 audio-visual combinations. In the auditory unimodal condition (n=5) the screen remained black. In the visual unimodal condition (n=5) a silent lip-movement was seen. Twelve 35-item lists, each comprising every token, randomly ordered, were constructed. A videotape was constructed from this material and participants in this experiment each viewed a total of 5 lists (total number of trials was 35 x 5 = 175 per participant) in the experimental condition.

Participants, who were screened for (self-reported) normal or corrected vision and who all reported normal hearing, were tested individually in a quiet lab. The videotape was viewed on a 22" colour monitor, with audio level set to 60dB. Viewing distance was about 50 cm. Participants were required to identify the spoken token by repeating 'what you think the speaker is saying'. For illustration, examples included a range of single and bi-consonants in the vocalic context of the study. For vision only, participants were asked 'what is she whispering?'

Behavior was monitored by video and audio recorder. Following ten unscored practice trials, experimental series were run in a single session, with a short break. All responses for WS participants were transcribed phonetically by the experimenter (MB), and checked independently by a second transcriber.

3 RESULTS

3.1 Unimodal presentations

The first analysis explored whether WS showed anomalous processing of auditory and of visual tokens presented unimodally. The response measure was accuracy of report (out of 5 for each token - means for each subject for each condition were therefore out of 25).

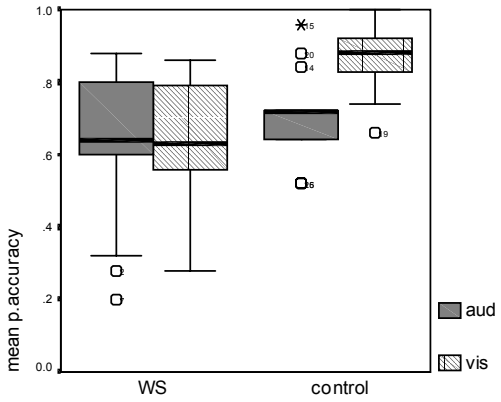

Figure 1: Accuracy of unimodal identification. Boxplots show medians of individual means for consonant identification (dark horizontal), boxes the interquartile range, whiskers the high-low range, excluding outliers (> 2SD). These are marked individually. The group difference for vision- alone is significant at p< 0.01.

This was a repeated measures analysis (SPSS GLM procedure) in which the within-subject (repeated measures) factors were modality and place of articulation. The between-subjects factor was experimental group, and age was entered as a covariate. There was a significant group x modality interaction ($F(1,23)= 4.76$, $p < 0.05$). WS were inferior to controls at visual, but not at auditory identification (post-hoc comparison , $p< 0.01$).

Modality affected place of articulation scores (modality x place interaction). $F(4,92) = 5.74$, $p < 0.01$), an effect which was not further moderated by experimental group or age. Audition was superior for discriminating the alveolar consonant /d/, vision was better for more anterior places of articulation (labial, labio-dental).

WS were less accurate than controls at discriminating visual, but not auditory tokens, while the pattern of performance (alveolars and velars better by ear, labials better by eye) was not distinguished by group.

3.2 Bimodal presentations

The next analysis (SPSS-GLM) explored bimodal responses. The dependent variable was mean probability of auditory identification (out of 5). The within-subject factors were: auditory place of articulation (5 levels), and visual place of articulation (5 levels). The between-subjects factor was again experimental status, covarying for age.

The only two significant (within-subjects) effects were: auditory x visual interaction [$F(16,368) = 11.42$, $p < 0.001$] and the auditory x visual x group interaction [$F(16,368) = 5.79$, $p<0.001$]. These are summarized in figure 2. This shows that where audition and vision were congruent, peak auditory accuracy was obtained. However, both this effect, and the converse (lowering of auditory accuracy by incongruent visual tokens) appear reduced in WS compared with controls.

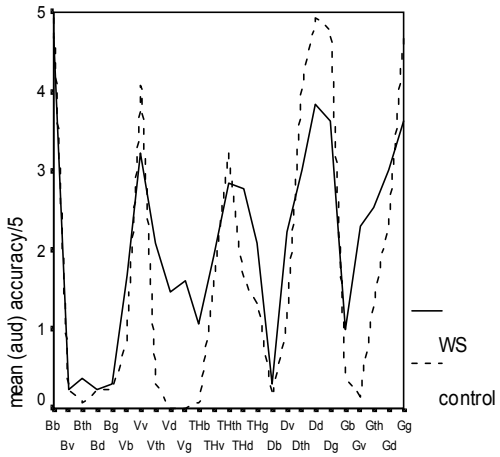

Figure 2 : bimodal performance, WS and controls compared. WS show reduced accuracy for congruent consonants and reduced influence of vision for some auditory tokens (eg /v/).

3.3 Are older WS like young controls?

A further analysis contrasted visual, audiovisual (congruent) and visual responses for older WS (n = 7, >16 years) and younger controls (n = 7, < 16 years). In this analysis the within-subject factors were modality (3 levels), place (5 levels). Despite the small numbers, there was a significant modality x group interaction ($F(2,24) = 8.13$, $p< 0.01$), moderated by place of articulation ($F(8,96) = 4.83$, $p< 0.001$)[1].

Posthoc comparisons confirmed that WS were poorer at vision-only than younger controls. In comparison with controls, visual /d/ and /g/ were particularly poorly discriminated ($p < 0.02$). They also confirmed that younger controls were poorer at auditory /v/ discrimination than other places of articulation ($p < 0.05$), but that this contrast was not significant for older WS. There were no significant differences in bimodal performance.

4 DISCUSSION

In this new syllable identification task, place of articulation affected report systematically for vision and for audition, with audition generally more accurate for speech articulated in more posterior parts of the vocal cavity, vision for consonants produced more anteriorly. There was clear evidence of integration in bimodal report. Both congruent and incongruent tokens showed marked effects. For example, the 'classic' McGurk paradigm [12], using auditory /b/ with visual /g/ (token Bg in figure 3), generated fewer than 10% 'b' responses in both tested groups.

4.1 Unimodal Processing – group effects

People with WS were impaired at identifying **visual** speech tokens, despite showing no significant differences from age-matched controls in auditory identification abilities. Nor was their pattern of visual speech identification simply developmentally delayed, since a comparison of young controls (<16 years) with older WS participants still showed significant differences in visual identification. While even the young controls could make use of visual information to distinguish /d/ and /g/ in this display, WS participants were poor at this, which contributed to their overall lower score.

4.2 Integration

Since WS were poorer at visual discrimination, vision should have a reduced effect on bimodal report, and so it proved. Nevertheless, there was extensive integration of vision and audition in WS, and young controls and older WS did not differ in this regard. These data cannot confirm failure of audiovisual integration in WS. The question cannot be closed, however. A strong test of the 'failure to integrate' hypothesis would match WS with other

[1] Only significant effects with calculated power value of > 0.85 are reported. This vitiates, to some extent, the use of small sample sizes and the interpretation of higher order interactions.

controls on the basis of similar visual identification accuracy, and explore the outcome on a case-by-case basis (see Massaro, [11] Chapter 5, for the rationale for individual psychometric analysis).

While an integration deficit in WS may be proposed for higher-level processes such as lexical and semantic-syntactic processing in online speech analysis [13,17], there is no clear indication from these data that cross-modal processing at the level of phoneme identification is defective in WS. Such a case has been made for other discrepant individuals and groups, including people with autism[5,6].

4.3 Accounting for the visual deficit in WS

Although people with WS are often observed to be socially sensitive and great 'lookers at faces' [3,14], their face processing can be shown to be compromised in various ways [7]. We have demonstrated that their ability to identify speech from facial gesture is anomalous compared with age-matched controls. What mechanisms may account for this? Facial configuration processing relating mouthshape to the position of tongue, teeth and lips may be required for visual consonant identification. In support of this, within the WS group the only significant correlation of any psychometric variable with visual speechreading (partialled for auditory performance) was with nonverbal tests of ability (BAS nonverbal test, and BAS nonverbal pattern construction subtest). /v/ accuracy correlated significantly with these test-scores (Spearman's rho = 0.67, $p < 0.02$).

WS may also have visual movement perceptual disorders. It has been established that the dorsal and ventral visual systems develop at different rates, and that dorsal system development, which is sensitive to visual movement, lags in WS [1,2]. This is a feasible neurophysiological locus for developmental problems related to a range of visual abilities in WS and may relate to the visual processing circuitry identified for speechreading [4].

5 REFERENCES

1. Atkinson J. Early visual development: differential functioning of parvocellular and magnocellular pathways. Eye 1992; **6**: 129-135
2. Atkinson J, King J, Braddick O, Nokes L, Anker S, Braddick. A specific deficit of dorsal stream function in Williams syndrome. NeuroReport 1997; **8**: 1919-1922.
3. Baron-Cohen S, Campbell R, Walker J, Karmiloff-Smith A. Are children with autism blind to the mentalistic significance of the eyes*?* British J. Developmental Psychology 1996; **13**: 379-398
4. Campbell R, Calvert G, Brammer M, MacSweeney M, Surguladze S, McGuire PK, Woll S, Williams S, Amaro E & David AS Activation In Auditory Cortex by Speechreading in hearing people Proceedings AVSP99
5. De Gelder B, Vroomen J, Bachoud-Levi A. Impaired speechreading in prosopagnosia. In R.Campbell, BJ Dodd, D Burnham (Eds)Hearing by Eye II, Psychology Press, Hove 1998
6. De Gelder B, Vroomen J, van der Heide L. Face recognition and lip-reading in autism. European J. Cognitive Psychology 1991; **31**: 69-86.
7. Deruelle C, Mancini J, Livet MO, Cassé-Perot C, de Schonen S. Processing of faces in children with Williams syndrome. Brain and Cognition 1999; **41**: 276-298.
8. Jordan TR, Bevan KM. Seeing and hearing rotated faces J. Experimental Psychology: HPP1997; **23**: 388-403.
9. Karmiloff-Smith A, Klima E, Bellugi U, Grant J, Baron-Cohen S. Is there a social module? Journal of Cognitive Neuroscience 1995; **7**: 196-208.
10. Massaro DW. Speech Perception by Ear and Eye Hillsdale NJ, Lawrence Erlbaum 1987.
11. Massaro DW. Perceiving talking faces. Cambridge Mass. MIT Press, 1998.
12. McGurk H, MacDonald JW. Hearing lips and seeing voices. Nature 1976; **264:** 746-748.
13. Mervis CB, Morris, CA, Bertrand J, Robinson B. Williams syndrome In H Tager-Flusberg (Ed) Neurodevelopmental Disorders. MIT Press, Cambridge, 1999; 65-110.
14. Reilly J, Klima ES, Bellugi U. Once more with feeling: Development and Psychopathology; 1990; **2**: 367-391.
15. Rosenblum LD, Johnson JA, Saldaña HM. Point-light facial displays enhance comprehension of speech in noise. J. Speech and Hearing Research, 1996; **39**: 1159-1170.
16. Rosenblum LD, Yakel DA, Green K Face and mouth inversion effects on visual and audiovisual speech perception J. Experimental Psychology: HPP;2000; **26**: 806-819
17. Tyler LK, Karmiloff-Smith A, Voice KL, Stevens T, Grant J, Udwin O, Davies M, Howlin P. Do individuals with Williams syndrome have bizarre semantics? Cortex 1997; **33:** 515-527.
18. Williams JCP, Barratt-Boyes BG, Lowe JB. Supravalvular aortic stenosis. Circulation 1961; **14**: 1311-1318