THE JOURNAL OF
CHILD PSYCHOLOGY AND PSYCHIATRY

Journal of Child Psychology and Psychiatry 49:9 (2008), pp 995–1000 doi:10.1111/j.1469-7610.2008.01902.x

Audiovisual speech integration in pervasive developmental disorder: evidence from event-related potentials

Maurice J.C.M. Magnée,[1,2] Beatrice de Gelder,[2,4] Herman van Engeland,[1] and Chantal Kemner[1,3]

[1]Rudolf Magnus Institute of Neuroscience, Department of Child and Adolescent Psychiatry, University Medical Center, Utrecht, The Netherlands; [2]Laboratory of Cognitive and Affective Neuroscience, Tilburg University, The Netherlands; [3]Section Biological Developmental Psychology, Faculty of Psychology, Maastricht University, The Netherlands; [4]Martinos Center for Biomedical Imaging, Massachusetts General Hospital and Harvard Medical School, Charlestown, MA, USA

Background: Integration of information from multiple sensory sources is an important prerequisite for successful social behavior, especially during face-to-face conversation. It has been suggested that communicative impairments among individuals with pervasive developmental disorders (PDD) might be caused by an inability to integrate synchronously presented visual and auditory cues. **Method:** We investigated audiovisual integration of speech stimuli among a group of high-functioning adult PDD individuals and age- and IQ-matched controls using electroencephalography, measuring both early pre-phonological, as well as late phonologically driven integration. **Results:** Pre-phonological AV interactions are intact, while AV interactions corresponding to more complex phonological processes are impaired in individuals with PDD. **Conclusions:** The present findings argue for a pattern of impairments on tasks related to complex audiovisual integration combined with relative sparing of low-level integrational abilities. This combination may very well contribute to the communicative disabilities which are typical for the disorder. **Keywords:** Multisensory integration, language and communication, autism, EEG, visual, auditory.

One of the major characteristics of pervasive developmental disorder (PDD) is a qualitative impairment in communicative abilities. This is manifested, amongst other things, by a delay in, or lack of, development of spoken language, deficits in the ability to initiate or sustain a conversation, and stereotyped or idiosyncratic use of language (American Psychiatric Association, 1994). The neurofunctional deficits underlying these problems in communication remain unclear. So far the majority of the studies investigating communicative deficits in PDD have concentrated on processing deficits in one modality only. However, studies among typically developing individuals show that integration of information across sensory systems also plays an important role in communication, especially during face-to-face conversation.

Difficulties with audiovisual (AV) integration have been suggested in the literature as an important problem in PDD (for review, see Iarocci & McDonald, 2006). Recent evidence, however, implies that the integration of low-level stimuli, such as bleeps and flashes, is normal in PDD (van der Smagt, van Engeland, & Kemner, 2007). Research on AV perception of higher-order stimuli such as emotions and speech, however, is still scarce. Behavioral studies on bimodal speech recognition in PDD brought conflicting results, as some report deficits (de Gelder, Vroomen, & van der Heide, 1991), while other groups argued that individuals with PDD show normal AV integration of speech stimuli (Massaro & Bosseler, 2003; Williams, Massaro, Peel, Bosseler, & Suddendorf, 2004). However, with standard behavioral methods it is difficult to rule out compensation strategies. Therefore, finding intact integration at the behavioral level does not necessarily imply that underlying neural mechanisms are intact. Investigating the temporal dynamics of AV speech integration using the high temporal resolution of ERPs may elucidate whether abnormalities in integrational abilities contribute to the observed impairments in communication in PDD.

In the present study we investigated AV integration of naturally occurring speech tokens in high-functioning adult males with PDD and age- and IQ-matched controls. Specifically, we were interested in interactions between auditory and visual speech perception both at low- and higher-order levels. Our experimental paradigm involved the presentation of a video of a woman's face producing the utterances /aba/ and /ada/. Incongruent AV pairs were created by dubbing auditory /aba/ with visual /ada/ and vice versa. By comparing both unimodal stimulus conditions to the bimodal condition, we were able to measure the influence of visual speech on early peaks in the auditory ERP (AEP), the N1 and P2. In line with previous research, we hypothesized that AEPs to AV speech stimuli would show reduced

Conflict of interest statement: No conflicts declared.

Published by Blackwell Publishing, 9600 Garsington Road, Oxford OX4 2DQ, UK and 350 Main Street, Malden, MA 02148, USA

amplitudes and temporal facilitation compared to both unimodal conditions, indicating low-level integration (van Wassenhove, Grant, & Poeppel, 2005). Furthermore, we hypothesized that this integration is intact in individuals with PDD. Additionally, the difference in ERP activity to congruent and incongruent AV speech was used to measure higher-order integrative processes of AV speech (Klucharev et al., 2003). We hypothesized that differences in late ERP components to these stimuli reflect detection of incongruency of phonological features of the stimuli, and we expected this higher-order integration to be abnormal in individuals with PDD.

Method

Participants

Twelve high-functioning, medication-free, adult males with PDD (average age 21.1, SD 4.0; one left-handed) and 13 healthy adult males (average age 23.0, SD 2.9; one left-handed) participated in the study. All individuals were administered the Wechsler Adult Intelligence Scale, Dutch edition (WAIS-III-NL). Mean age and total IQ scores were similar for individuals with PDD (TIQ 119.4, SD 10.6) and individuals from the control group (TIQ 127, SD 14.4). All individuals in the control group were free of neurological and psychiatric history, and familial history of psychiatric disorders.

All diagnoses of PDD (either autistic disorder or Asperger syndrome, no comorbidities) were based on DSM-IV criteria and were made by a child psychiatrist. Additionally, all patients were administered the Autism Diagnostic Observation Schedule (ADOS; Lord et al., 1989) by a trained rater, and their parents were informants on the Autism Diagnostic Interview Revised (ADI-R; Lord, Rutter, & Le Couteur, 1994). Six individuals with PDD met full ADI-R and ADOS criteria for autism or autism spectrum disorder. We were not able to acquire ADOS scores for two patients, but both fulfilled ADI-R criteria. Three individuals met criteria on the ADOS and scored one or two points below cut-off on one scale of the ADI-R. One individual scored one point below cut-off on both ADI-R (stereotyped behavior) and ADOS (social behavior) criteria. Mean ADOS scores in the PDD group were 4.1 (SD 1.8) for the subscale 'Communication' (cut-off 2), and 8.4 (SD 3.6) for 'Social behavior' (cut-off 4). Mean ADI scores were 19.2 (SD 3.6) for 'Social behavior' (cut-off 10), 15.6 (SD 4.9) for 'Communication' (cut-off 8), 5.6 (SD 3.3) for 'Repetitive behaviors' (cut-off 3), and 2.7 (SD 1.1) for 'Age of onset' (cut-off 1).

All participants had normal or corrected to normal vision. They were all paid for their participation. Written informed consent was obtained for each participant before the session, according to the Declaration of Helsinki (2000). Approval of the medical ethics committee of the University Medical Center Utrecht was obtained prior to the study.

Stimuli and procedure

Stimulus material consisted of a video of a woman's face producing the utterances /aba/ and /ada/. Each utterance included a total of 12 frames, with each frame being presented for 85 ms (total duration 1020 ms). Articulatory movements started 95 ms prior to audio onset as is the case in naturally produced speech, and audio onset started 350 ms after visual onset. Duration of the auditory stimuli was 540 ms. Incongruent AV pairs were created by dubbing auditory /aba/ with visual /ada/ and vice versa. Incongruency between auditory and visual stimuli became noticeable after six frames (510 ms after visual onset; see Figure 1), at which point differences in articulatory and acoustic phonetics of the consonants (/b/ vs. /d/) could be distinguished. Congruent and incongruent stimuli consisted of the same acoustic and visual components. Incongruent trails were clearly conflicting, and auditory and visual components were not perceptually fused. Catch trials were included in order to control for visual attention. Here, a small blue dot was positioned on the nose of the speaker during the 7th frame. Participants were instructed to push a designated button during these trials. All catch trials were excluded from further analyses.

Audiovisual, auditory (A) and visual (V) trials were presented in separate blocks. AV blocks consisted of 90 congruent, 90 incongruent and 20 catch trials. V and A

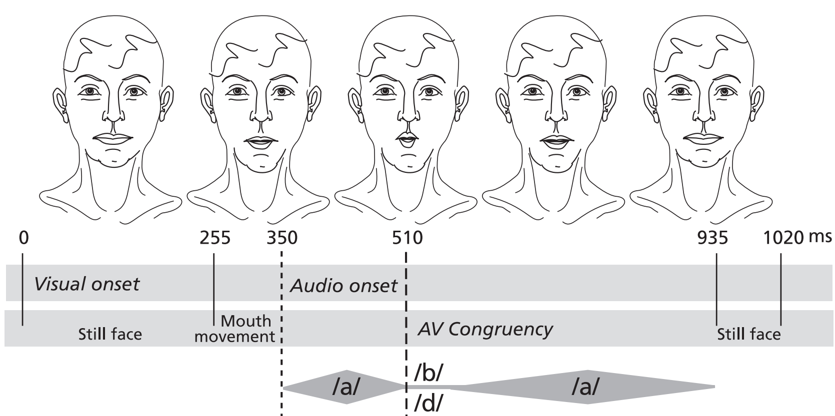

Figure 1 Time course of an AV speech trial. Each trial started with the presentation of a still face. Mouth movement started at 255 ms after visual onset and preceded audio onset by about 95 ms; AV congruency became noticeable 510 ms after visual onset. In the V and A conditions the time course was similar to AV trials

blocks consisted of 108 experimental and 12 catch trials. Catch trials were included in A and V blocks as well. Catch trials in A blocks consisted of an auditory stimulus and the visual presentation of a similar blue dot positioned at the centre of the screen during the 7th frame.

The size of the portraits was 10 cm height × 6.5 cm width, which at the mean viewing distance of 80 cm corresponds to a visual angle of 7.2°× 4.7. The mean luminance of the pictures was 22 cd/m^2 on a 2.5 cd/m^2 background. The mean level for sound was 63 dB(a) delivered over one loudspeaker placed directly below the screen. Participants were seated in a soundproof experimental chamber. They were instructed to judge what was said, by pushing one out of three designated buttons on a response box. For AV conditions this meant individuals had to respond to the voice while looking at the face. In order to avoid any response-related components in the ongoing ERP signal, they were furthermore instructed not to respond until after offset of the visual stimulus. Each stimulus was preceded by the presentation of a central fixation cross during 500 ms. In addition, an intertrial interval was chosen randomly between 750 and 1500 ms, measured from the participant's response.

Recordings

EEGs were recorded from 48 locations using standard Ag/AgCl pin-type active electrodes (BIOSEMI, Amsterdam, The Netherlands) mounted in an elastic cap, referenced to an additional active electrode (Common Mode Sense) during recording. EEG signals were band-pass filtered (.1–30 Hz) at a sample rate of 512 Hz and offline referenced to an average reference. Horizontal and vertical EOGs were measured for offline correction (Gratton & Coles).

The raw data were segmented into 1600-ms epochs for visual, auditory and audiovisual categories separately, including a 50-ms prestimulus baseline. After EOG correction, epochs with amplitudes exceeding ±100 μV at any channel were automatically rejected. Lowest allowed activity was 3μV/200 ms, and the maximal allowed voltage step per sampling point was 50 μV. ERPs were averaged separately for all four categories (V, A, AV congruent, and AV incongruent).

Data analysis

Peak latencies and amplitudes of AEPs were measured at frontal-central electrodes (FC1, FC2, Cz, CPz), based on previous studies showing maximal amplitudes on these sites (e.g., Besle et al., 2004). N1 and P2 latencies and amplitudes were determined as the mean of the individual peaks over the four electrodes, and measured as maximal negative peaks in the time window 80–160 ms after auditory onset, P2 as maximal positive peaks in the time window 180–260 ms after auditory onset, for V, A, and AV conditions.

Analyses of late activation were based on pronounced differences between congruent and incongruent AV trials (non-blinded). Since the literature provides little guidance on the expected location of the effects, mean activity of subsequent 50-ms time-windows was calculated around the located maximum difference between both conditions for 800–1300 ms poststimulus, corresponding to 300–800 ms after onset of incongruency. Specifically, this resulted in pooled ERP activity (μV) at left frontal sites (Fp1, AF3, F3, F7, FC5), right frontal sites (Fp2, AF4, F4, F8, FC6) and central-parietal sites (C1, Cz, C2, CP1, CPz, CP2, P1, Pz, P2, PO3, POz, PO4; see Figure 3).

Using a multivariate analysis of variance (MANOVA) procedure for repeated measures, we first tested whether N1 and P2 latencies and amplitudes differed in response to bimodal stimuli (AV) compared to the sum of ERP signals obtained in unimodal conditions (AV – (A + V)). This method was used by Giard and Peronnet (1999) for determining whether AV responses could be merely accounted for by adding auditory and visual evoked potentials. For these comparisons MANOVA analyses consisted of one between-subjects factor Group (PDD vs. control group) and the within-subjects factor Modality (AV vs. A + V).

We tested AV congruency effects by comparing AV congruent and incongruent speech dubbings. Both left and right frontal pools were analyzed over 50-ms time windows, consisting of one between-subjects factor Group (PDD vs. control group) and the within-subjects factors Hemisphere (left vs. right) and Congruency (congruent vs. incongruent). As frontal and parietal areas are thought to serve different functions in AV integration (Miller & D'Esposito, 2005), mean pooled central-parietal ERP activity was analyzed separately, with Group as between-subjects factor (PDD vs. control group) and Congruency as within-subjects factor (congruent vs. incongruent).

Results

Behavioral results

There were no group differences in detection of catch trials. Both groups performed almost perfectly on the forced categorization to judge what was said (aba vs. ada) during V, A, and AV trials.

Electrophysiological data

N1 and P2 amplitudes and latencies. Significant temporal facilitation of the N1 was found during AV speech, $F(1,23) = 37.02$, $p < .0001$. Mean N1 latency in the AV condition was found at 147 ms, in the (A + V) condition at 167 ms. There were no significant group interaction effects. A comparable effect was also observed on P2 latencies, $F(1,23) = 16.61$, $p < .0001$, for both groups. Mean P2 latency in the AV condition was elicited at 225 ms, in the (A + V) condition at 238 ms. Significant N1 amplitude reduction was found in the AV condition, compared to the sum of A and V conditions, $F(1,23) = 8.30$, $p < .01$, and this effect was present in both groups. Mean N1 amplitude in the AV condition was −1.67 μV, in the (A + V) condition −2.59 μV. No significant effect of Modality was observed on P2 amplitudes, with no significant Modality × Group interaction. Mean P2 amplitude in the AV condition was 1.79 μV, in the (A + V) condition 1.57 μV (Figure 2).

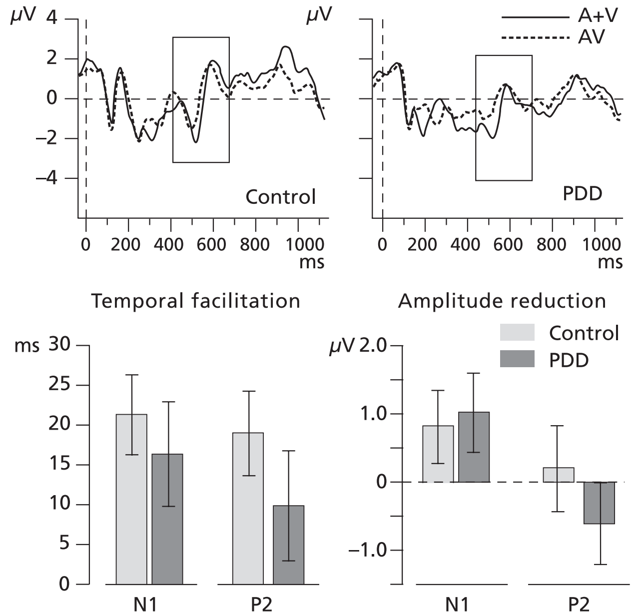

Figure 2 Above: Frontal-central auditory evoked potentials in response to AV stimuli, and the sum of both unimodal (A + V) stimuli in the control and PDD group. N1 and P2 peaks are shown within the frame. Below: Temporal facilitation and amplitude reduction for AV conditions compared to the sum of unimodal conditions

AV congruency. We found for both frontal and central-parietal pools a significant Congruency × Group interaction in the four consecutive time windows between 1100 and 1300 ms poststimulus onset, corresponding to 600–800 ms after onset of incongruency. There were no significant hemispheric differences. Specifically, regarding these four time windows Congruency × Group values of left and right frontal sites revealed statistical values varying between $F(1,23) = 5.87$, $p < .05$, and $F(1,23) = 7.48$, $p < .05$. For the analyses of central-parietal sites, Congruency × Group values varied between $F(1,23) = 4.71$, $p < .05$, and $F(1,23) = 8.16$, $p < .01$.

When testing the effects of Congruency for both groups separately, significant effects were found in the control group in these time windows (see Figure 3). With respect to frontal electrode pools, statistical values varied between $F(1,12) = 10.00$, $p < .01$, and $F(1,12) = 15.06$, $p < .01$. On average, this was reflected in more frontal negativity in the AV incongruent (−2.8 μV) compared to the AV congruent condition (−1.6 μV). None of these analyses revealed significant effects of Congruency in the PDD group, as we found no differences between AV incongruent (−1.6 μV) and AV congruent speech (−1.7 μV). Significant Congruency effects were found in the control group in the same time windows for the central-parietal electrode pool as well, with statistical values varying between $F(1,12) = 9.91$, $p < .01$, and $F(1,12) = 19.25$, $p < .001$. Here, AV incongruent speech elicited larger positivity (2.2 μV) compared to AV congruent speech (1.3 μV). Again, none of these tests revealed significant effects in the PDD group (incongruent 1.5 μV compared to congruent 1.6 μV).

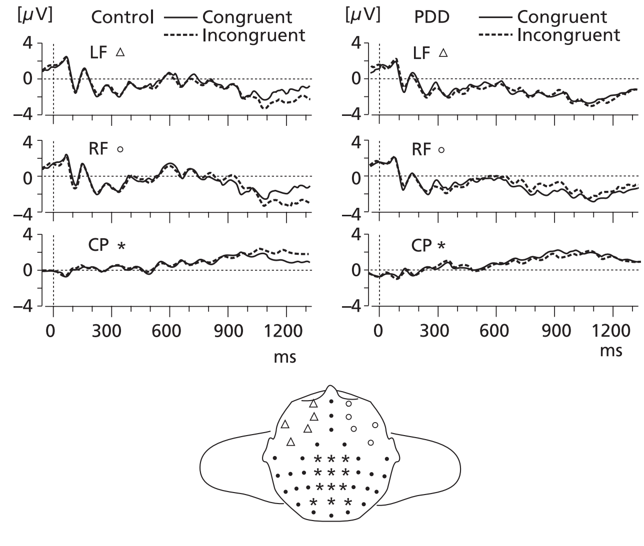

Figure 3 Mean pooled ERP activity at left frontal (LF), right frontal (RF), and central-parietal (CP) electrode sites, representing congruent and incongruent AV speech of both control group and PDD group. Differences between the two conditions were noticed at 1100 ms (590 ms after incongruency) in the control group only

Discussion

Our goal was to investigate the temporal dynamics of AV speech integration in individuals with PDD and matched controls. The first part of our study focused on two early AEP components, the N1 and P2. The results indicated reduced amplitudes (N1) and temporal facilitation (N1 and P2) of AEPs to AV speech compared to responses to unimodal stimuli. This phenomenon was previously observed in healthy individuals and is thought to indicate that the presence of visual speech information facilitates auditory processing at this early level of processing (van Wassenhove et al., 2005). Recently, it has been shown that this low-level AV interaction is not speech-specific, but merely related to the degree in which the visual signal predicts the auditory stimulus (Stekelenburg & Vroomen, 2007). Here we show that this signature effect is also observed in PDD patients, providing additional evidence for normal low-level integration in PDD (van der Smagt et al., 2007).

Secondly, we were interested in higher-order integrative processes of AV speech. We found clear evidence of an AV congruency effect in the control group, consisting of late (starting at 500 ms after onset) bilateral frontal negativity and central-parietal positivity to incongruent compared to congruent AV speech. However, this congruency effect was not present in PDD patients. Studies of typically developing individuals have related differences between AV congruent and incongruent speech to higher-order integration of phonological information in bilateral brain regions (Klucharev et al., 2003). Note that in this study and in the present one no hemispherical differences were found, suggesting that AV speech integration depends on the functioning of both hemispheres. The finding of a lack of this congruency effect in individuals with PDD is consistent with a view of PDD as a disorder of complex information processing (Williams et al., 2006). According to this view, the most affected domains of processing are those that place the highest demands on information-processing capacity, while low-level abilities may be intact. In this sense, atypicalities in processing have more to do with the mere complexity of the process rather than the type of processing that is being done. For instance, besides impairments in AV speech interactions, anomalous integrated processing of AV emotional information has also been found (Hall et al., 2003). This argues for a pattern of impairments on tasks that put high demands on integration of information and relative sparing on more low-level tasks.

An important question is how this AV congruency effect relates to cognitive functioning. The observed effect shares temporal characteristics with processes that are specifically related to processing of linguistic information. The most prominent ERP peak in this respect described in the literature in the latency window of 500 to 800 ms is the P600, which has traditionally been associated with several kinds of syntactic violations (Osterhout, Holcomb, & Swinney, 1994). Recent studies related P600 effects to the integration of semantic, syntactic and pragmatic information (Eckstein & Friederici, 2005), and to a process that monitors the correctness of the percept (van Herten, Kolk, & Chwilla, 2005). This interpretation suggests that in individuals with PDD this monitoring of ambiguous percepts may not be functioning correctly.

Our study contributes to a better understanding of the problems in communication found in PDD, yet it also has some limitations. Firstly, we did not include individuals with related disorders like ADHD or schizophrenia, who may seemingly also have problems on perceptual processing but with a very different origin including pharmacological treatment. Secondly, we took great effort including a homogeneous group of adult individuals without intellectual disabilities as patients, which resulted in relatively small sample sizes, thereby reducing the statistical power of our effects. As high-functioning individuals with PDD tend to acquire compensational skills, they usually score low on diagnostic interviews such as the ADOS. Although these individuals scored well above ADOS cut-off values on group level and study results were based on group averages, the present results cannot be generalized to a younger population or to a group of lower-functioning individuals with PDD. Furthermore, one should not overlook the effect of attentional capacity, which can already modulate AV integration at the earliest stages of the process (Talsma, Doty, & Woldorff, 2007). In the present study attention was assured through the use of a concurrent speech recognition task, although slight differences between the two groups may still have confounded the results. The question whether atypicalities in AV integration in PDD arise solely from deficits in processes related to integrational abilities, or also from impairments in attentional capacity, therefore, remains to be answered.

Conclusion

In sum, we suggest that in PDD patients early non-linguistic AV interactions are intact, while AV interactions corresponding to more complex phonological processes are impaired. Given the importance of rapid multisensory integration for successful social behavior, deficits in the integration of phonological information may very well contribute to the communicative disabilities which are typical for the disorder.

Acknowledgements

This research was funded by an Innovational Research Incentives grant of the Netherlands Organization for Scientific Research (NWO; VIDI-scheme, 402-01-094) to Chantal Kemner.

Correspondence to

Maurice Magnée, Department of Child and Adolescent Psychiatry, University Medical Center Utrecht, B01.201, Heidelberglaan 100, 3584 CX Utrecht, The Netherlands; Tel: + 31 88 755 6026; Fax: + 31 88 755 5444; Email: M.J.C.M.Magnee@umcutrecht.nl

References

American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th edn). Washington, DC: American Psychiatric Association.

Besle, J., Fort, A., Delpuech, C., & Giard, M.H. (2004). Bimodal speech: Early suppressive visual effects in human auditory cortex. *European Journal of Neuroscience*, *20*, 2225–2234.

de Gelder, B., Vroomen, J., & van der Heide, L. (1991). Face recognition and lip-reading in autism. *European Journal of Cognitive Psychology*, *3*, 69–86.

Eckstein, K., & Friederici, A.D. (2005). Late interaction of syntactic and prosodic processes in sentence comprehension as revealed by ERPs. *Cognitive Brain Research*, *25*, 130–143.

Giard, M.H., & Peronnet, F. (1999). Auditory-visual integration during multimodal object recognition in humans: A behavioral and electrophysiological study. *Journal of Cognitive Neuroscience*, *11*, 473–490.

Hall, G.B.C., Szechtman, H., & Nahmias, C. (2003). Enhanced salience and emotion recognition in autism: A PET study. *American Journal of Psychiatry*, *160*, 1439–1441.

Herten, van M., Kolk, H.H., & Chwilla, D.J. (2005). An ERP study of P600 effects elicited by semantic anomalies. *Cognitive Brain Research*, *22*, 241–255.

Iarocci, G., & McDonald, J. (2006). Sensory integration and the perceptual experience of persons with autism. *Journal of Autism and Developmental Disorders*, *36*, 77–90.

Klucharev, V., Möttönen, R., & Sams, M. (2003). Electrophysiological indicators of phonetic and non-phonetic multisensory interactions during audiovisual speech perception. *Cognitive Brain Research*, *18*, 65–75.

Lord, C., Rutter, M., Goode, S., Heemsbergen, J., Jordan, H., Mawhood, L., & Schopler, E. (1989). Autism Diagnostic Observation Schedule: A standard observation of communicative and social behavior. *Journal of Autism and Developmental Disorders*, *19*, 185–212.

Lord, C., Rutter, M., & Le Couteur, A. (1994). Autism Diagnostic Interview-Revised: A revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. *Journal of Autism and Developmental Disorders*, *24*, 659–685.

Massaro, D.W., & Bosseler, A. (2003). Perceiving speech by ear and eye: Multimodal integration by children with autism. *Journal on Development and Learning Disorders*, *7*, 111–144.

Miller, L.M., & D'Esposito, M.D. (2005). Perceptual fusion and stimulus coincidence in the crossmodal integration of speech. *Journal of Neuroscience*, *25*, 5884–5893.

Osterhout, L., Holcomb, P.J., & Swinney, D.A. (1994). Brain potentials elicited by garden-path sentences: Evidence of the application of verb information during parsing. *Journal of Experimental Psychology, Learning, Memory, and Cognition*, *20*, 786–803.

Smagt, van der M.J., van Engeland, H., & Kemner, C. (2007). Can you see what is not there? Low-level auditory-visual integration in autism spectrum disorder. *Journal of Autism and Developmental Disorders*, *37*, 2014–2019.

Stekelenburg, J.J., & Vroomen, J. (2007). Neural correlates of multisensory integration of ecologically valid audiovisual events. *Journal of Cognitive Neuroscience*, *19*, 1964–1973.

Talsma, D., Doty, T.J., & Woldorff, M.G. (2007). Selective attention and audiovisual integration: Is attending to both modalities a prerequisite for early integration? *Cerebral Cortex*, *17*, 679–690.

Wassenhove, van V., Grant, K.W., & Poeppel, D. (2005). Visual speech speeds up the neural processing of auditory speech. *Proceedings of the National Academy of Sciences of the USA*, *102*, 1181–1186.

Williams, D.L., Goldstein, G., & Minshew, N.J. (2006). Neuropsychologic functioning in children with autism: Further evidence for disordered complex information-processing. *Child Neuropsychology*, *12*, 279–298.

Williams, J.H.G., Massaro, D.W., Peel, N.J., Bosseler, A., & Suddendorf, T. (2004). Visual-auditory integration during speech imitation in autism. *Research in Developmental Disabilities*, *25,* 559–575.

Manuscript accepted 8 January 2008