Autism Speaks
Author Manuscript
Accepted for publication in a peer-reviewed journal

Published in final edited form as:
Autism Res. 2012 February ; 5(1): 39–48. doi:10.1002/aur.231.

Audiovisual speech integration in autism spectrum disorder: ERP evidence for atypicalities in lexical-semantic processing

Odette Megnin[1], **Atlanta Flitton**[1], **Catherine Jones**[2], **Michelle de Haan**[3], **Torsten Baldeweg**[3], and **Tony Charman**[2]

[1]Behavioural and Brain Sciences Unit, UCL Institute of Child Health, 30 Guilford Street, London, WC1N 1EH, UK

[2]Centre for Research in Autism and Education, Institute of Education, 20 Bedford Way, London, WC1H 0AL, UK

[3]Developmental Cognitive Neuroscience Unit, UCL Institute of Child Health, 30 Guilford Street, London, WC1N 1EH, UK

Abstract

Lay Abstract—Language and communicative impairments are among the primary characteristics of autism spectrum disorders (ASD). Previous studies have examined auditory language processing in ASD. However, during face-to-face conversation, auditory and visual speech inputs provide complementary information, and little is known about audiovisual (AV) speech processing in ASD. It is possible to elucidate the neural correlates of AV integration by examining the effects of seeing the lip movements accompanying the speech (visual speech) on electrophysiological event-related potentials (ERP) to spoken words. Moreover, electrophysiological techniques have a high temporal resolution and thus enable us to track the time-course of spoken word processing in ASD and typical development (TD). The present study examined the ERP correlates of AV effects in three time windows that are indicative of hierarchical stages of word processing. We studied a group of TD adolescent boys (n=14) and a group of high-functioning boys with ASD (n=14). Significant group differences were found in AV integration of spoken words in the 200–300ms time window when spoken words start to be processed for meaning. These results suggest that the neural facilitation by visual speech of spoken word processing is reduced in individuals with ASD.

Scientific Abstract—In typically developing (TD) individuals, behavioural and event-related potential (ERP) studies suggest that audiovisual (AV) integration enables faster and more efficient processing of speech. However, little is known about AV speech processing in individuals with autism spectrum disorder (ASD). The present study examined ERP responses to spoken words to elucidate the effects of visual speech (the lip movements accompanying a spoken word) on the range of auditory speech processing stages from sound onset detection to semantic integration. The study also included an AV condition which paired spoken words with a dynamic scrambled face in order to highlight AV effects specific to visual speech. Fourteen adolescent boys with ASD (15–17 years old) and 14 age- and verbal IQ-matched TD boys participated. The ERP of the TD group showed a pattern and topography of AV interaction effects consistent with activity within the superior temporal plane, with two dissociable effects over fronto-central and centro-parietal regions. The posterior effect (200–300ms interval) was specifically sensitive to lip movements in TD boys, and no AV modulation was observed in this region for the ASD group. Moreover, the magnitude of the posterior AV effect to visual speech correlated inversely with ASD

Address for correspondence: Dr. Odette Megnin, Centre for Neuroscience in Education, Experimental Psychology, University of Cambridge, Downing Street, Cambridge, CB2 3EB, om252@cam.ac.uk, Tel: +44 (0)1223 767522, Fax: +44 (0)1223 767602.

Autism Speaks Author Manuscript
Autism Speaks Author Manuscript

symptomatology. In addition, the ASD boys showed an unexpected effect (P2 time window) over the frontal-central region (pooled electrodes F3, Fz, F4, FC1, FC2, FC3, FC4) which was sensitive to scrambled face stimuli. These results suggest that the neural networks facilitating processing of spoken words by visual speech are altered in individuals with ASD.

Keywords

Auditory; ASD; ERP; Language; Multisensory; Visual

Introduction

Behavioural and event-related potential (ERP) studies of typically developing individuals suggest that viewing a speakers' lip movements enables faster and more efficient processing of spoken syllables (e.g. Besle et al. 2004; Klucharev et al. 2003; Stekelenburg & Vroomen, 2007; van Wassenhove et al. 2005). However, little is known about the audiovisual (AV) integration of speech in individuals with autism spectrum disorder (ASD). This is important as language and communicative impairments are among the primary characteristics of autism and a lack of attention to faces is one of the first evident symptoms (Osterling & Dawson, 1994).

Bebko et al. (2006) found that children with ASD were not sensitive to the temporal synchrony of linguistic AV events, i.e. the ASD group did not preferentially look at a screen showing lip movements which were synchronised to the sound, compared to a screen where the lip movements were offset from the sound by a delay. In addition, group differences have been found for AV illusions. For instance, the McGurk effect (McGurk & MacDonald, 1976), which describes a bimodal illusion whereby an incongruent auditory and visual token (e.g. auditory "ba" and the lip movements "ga") produce an illusory percept (e.g. "da"), provides behavioural evidence for an automatic and mandatory effect of visual speech (i.e. information conveyed by lip movements) on auditory speech processing in typical development. However, adolescents with autism have been found to report fewer McGurk fusions than typically-developing (TD) children (de Gelder et al. 1991; Williams et al. 2004). Iarocci et al. (2010) found that children with autism showed less visual influence and more auditory influence on their bimodal speech perception, and this was largely due to significantly worse performance in the unimodal visual condition (lip reading). However, Smith and Bennetto (2007) found that although their ASD group were worse than the TD group at hearing speech-in-noise in an AV condition there was evidence for some facilitation with visual speech cues, i.e. the ASD group showed better relative performance in the AV condition than in the auditory-only condition.

Electrophysiological studies using ERP are ideally suited to the examination of speech and language processing due to their temporal resolution allowing examination of individual processing stages. In the present study electrophysiological AV effects were examined in three time windows indicative of consecutive stages of spoken word processing: (a) word onset detection as indexed by the N1 component (see Čeponiené et al. 2005; Sanders et al. 2002); (b) the transition from phonetic to lexical-semantic analysis as indexed by the P2 component (see Bentin et al. 1985; Pulvermüller et al. 2009); and (c) semantic integration as indexed by modulations of the N4 component (Domalski et al. 1991; Kutas & Hillyard, 1980; Rugg, 1985). Although behavioural studies have suggested atypicalities in AV speech perception in ASD very few ERP studies have been reported. Magnée et al. (2008) examined AV integration of naturally occurring speech tokens (/aba/ and /ada/) in high-functioning adult males with ASD and found no evidence for group differences in the early time window ERP. Both the ASD and TD group showed a temporal facilitation of the N1 and P2 and significant N1 amplitude attenuation during AV speech suggesting that the initial

Autism Speaks Author Manuscript Autism Speaks Author Manuscript

Autism Speaks Author Manuscript

Autism Speaks Author Manuscript

syllable processing stages, e.g. sound onset detection were not impaired in ASD. However, that study did find group differences in later phase processing, with only the TD group showing evidence for an AV congruency effect consisting of a late (>500ms post-onset) bilateral frontal negativity and central-parietal positivity to incongruent compared to congruent AV speech. However, the use of speech syllables as opposed to whole spoken words (with semantic content) meant that further investigation of this later-phase group difference in AV integration was not possible.

Subtle abnormalities in semantic processing of speech have been found in ASD, even where language skills are within the normal range (e.g. Harris et al. 2006; though see Norbury, 2005). Individuals with ASD have been found to show difficulties in using context to predict meaning, for instance, failing to use sentence context to determine the correct pronunciation of homographs (Frith & Snowling, 1983; Snowling & Frith, 1986; Happé, 1997). Furthermore, a number of ERP studies have provided evidence for reduced semantic integration in ASD, as reflected by differences in modulations of the N4 component. For instance, Ring et al. (2007) found that TD controls showed larger N4 amplitude to semantically incongruent relative to congruent sentence endings. However, an ASD group showed no significant differences in N4 amplitude between congruent and incongruent endings. This group difference in N4 amplitude modulation has also been replicated on a word priming task, with only TD controls showing enhanced N4 amplitude response to out-of-category words in relation to in-category words (Dunn et al. 1999; Dunn & Bates, 2005). Clearly, a thorough investigation of the neural signature of AV integration in ASD should include examination of the processing of semantic information. This study explored AV speech processing up to the level of semantic integration in ASD. Instead of using a congruency paradigm, cross-modal semantic integration was examined by measuring the effects of lip movements on modulating the amplitude of the N4 response to the spoken word. The hypothesis is that group differences in AV effects will be confined to higher levels of word processing.

Finally, in addition to examining the effects of accompanying lip movements on auditory processing in TD and ASD groups, an audiovisual scrambled face (AVS) condition was included in order to elucidate AV effects specific to phonetically informative lip movements over and above dynamic (temporal) visual cues in the scrambled face stimuli. This was adopted as an alternative to including incongruent speech, as misleading lip movements may result in ‘wait and see’ processing rather than a predictive integrative strategy, i.e. using lip movements to predict the identity of the auditory word (Kutas & Federmeier, 2000). Inclusion of a non-face AV control condition was also particularly important in light of specific atypical responses to faces in ASD (e.g. McPartland et al. 2004). Differences between responses to the AV condition with a face and to the AV condition with a scrambled face will elucidate potential group differences in the contribution of phonetic visual speech to auditory word processing.

Materials And Methods

Participants

Fourteen TD and 14 high-functioning adolescent boys with ASD participated in this study (see Table 1 for participant characteristics). Participants were paid £10 an hour to participate in this study. The study was approved by the Institute of Child Health/Great Ormond Street Hospital Research Ethics Committee (06/Q0508/113) and all participants gave informed consent in accordance with the human subject research protocol. All participants were free of neurological disease, had normal hearing and normal or corrected to normal sight. All of the participants with ASD, and 9 of the TD participants were recruited from the Special Needs and Autism Project (SNAP; Baird et al. 2006). For the ASD cohort, consensus

Autism Speaks Author Manuscript

Autism Speaks Author Manuscript

clinical ICD-10 diagnoses were made by three experienced clinicians using information from the ADI-R (Lord et al. 1994) and ADOS-G (Lord et al. 2000) as well as IQ, language and adaptive behaviour measures (see Baird et al. 2006 for details). Six participants in the ASD sample had a diagnosis of childhood autism, while the remaining eight met criteria for 'other ASD' (Pervasive Developmental Disorder). The TD participants were recruited from local mainstream schools. The Social Communication Questionnaire (SCQ; Rutter et al. 2003) was also collected from all participants. The SCQ is a 40-item parent-report questionnaire with each item rated 0 or 1 where 1 represents endorsement of each symptom of autism. Half the items rate current behaviour and half rate behaviour when the child was 4–5 years old. None of the participants in the TD group scored 15 or above which is the cut-off for ASD. Three participants in the ASD group scored below 15 on the SCQ, however, they were included as they had received a diagnosis from three experienced clinicians. IQ was measured with the Wechsler Abbreviated Scale of Intelligence-UK (WASI; Wechsler, 1999). There were no significant group differences found in age ($t(26)=0.06$, $p=0.9$) or VIQ ($t(26)=1.5$, $p=0.1$). However, TD participants did have a significantly higher PIQ than the ASD group ($t(26)=2.2$, $p=0.03$). Consequently in the analyses that follow PIQ was entered as a covariate. Finally, participants were required to pass a hearing screen using an Earscan 3 audiometer.

Stimuli

Monosyllabic spoken words (60 per condition) were presented in one of five conditions (Figure 1): auditory-only (A); visual-only with face (VF); audiovisual with face (AVF); visual-only with scrambled face (VS); and audiovisual with scrambled face (AVS). In the A condition participants heard the word and saw a blank screen. In the VF condition participants saw the lip movements of a word but heard no sound, and in the AVF condition participants heard the word and saw the congruent lip movements. Scrambled face control conditions were created by inverting the image and applying a mosaic effect. This effect divides the screen into squares and replaces the luminance in each square with the mean value, creating a pixellated appearance. This ensured equivalent luminance and dynamic activity but removed visual phonetic information as lip movements were not identifiable. For these control stimuli, participants saw the scrambled face and heard nothing (VS condition), or saw a scrambled face and heard a word (AVS condition). Spoken words were matched for word length and frequency of occurrence in verbal language (derived from the London-Lund Corpus of English Conversation by Brown, 1984) across conditions. The mean duration of the auditory words was 537ms (with words ranging between 200ms and 900ms). AV stimuli consisted of video clips of natural speech produced by four different actors (two male and two female). The videos were edited using Adobe Premiere Pro 2.0 with a digitization rate of 30 frames per second. Video images were cropped to display full-head views. AVF, VF, AVS, and VS stimuli began with a static image (for 500ms) to prevent the auditory evoked potential being contaminated by visual ERP responses to the face (for instance, the N170 component). The stimuli were temporally aligned so that the auditory onset was 1000ms from the start of the trial. The onset of mouth movements differed from word-to-word with initial mouth movement onset at a mean of 668ms (with onsets ranging from 533ms to 867ms). As stimuli consisted of naturally produced speech, lip movements always preceded auditory onset (by a mean of 332ms). In all conditions, the total duration of each stimulus trial was 2000ms.

Procedure

Participants were required to watch and/or listen to short videos of spoken words and perform a target detection task. 24 targets were presented in each of A, VF, AVF, and AVS conditions. The target consisted of a linguolabial trill (sound made by sticking the tongue between the lips and blowing, known colloquially as a 'raspberry' sound) with or without

Autism Speaks Author Manuscript

Autism Speaks Author Manuscript

accompanying visual information. Target stimuli followed the same time course as the spoken words. A target could not be presented in the VS condition as there would be no grounds on which to detect it with no sound and undecipherable lip movements. The inter-stimulus interval (ISI) was 1000ms between two consecutive experimental stimuli. Participants were seated in a dimmed, sound-attenuated room. Videos were displayed with images of 17.7° × 12.8° centred on a black background. Sounds were presented from a central speaker at approximately 70dB SPL. Subjects were presented with stimuli in a pseudorandom presentation order in four blocks (each block contained all conditions) with short breaks between blocks.

EEG recording

EEG recordings were obtained using the EasyCap (EASYCAP GmbH, Germany) with 41 sintered Ag-AgC1 ring electrodes arranged in accordance with the international 10-10 system: Fz, Cz, CPz, Pz, Oz, FP1, F3, F7, FC1, FC3, FC5, FT9, C3, T7, CP1, CP3, CP5, TP9, P3, P7, PO3, PO7, PO9, and their counterparts in the right hemiscalp. The EEG was recorded continuously (NeuroScan Systems, ACQUIRE 4.3; NeuroScan Labs, Sterling, VA) through Synamps AC coupled amplifiers (0.05 – 70Hz analogue bandwidth) with a sampling rate of 500Hz. Electrode impedance was measured at the beginning of each recording session and quantified as impedance levels of below 5kΩ. The online EEG recording was referenced to channel CPz, and the ground electrode was positioned on the central forehead (channel FPz). Horizontal and vertical electrooculograms (EOG) were recorded from bipolar electrodes placed on the left and right outer canthus (channels F9 and F10 – HEOG) and above and below the right eye (VEOG and channel FP2), for off-line artefact reduction.

Data analysis

ERP analysis—An ocular artefact reduction algorithm (Semlitsch et al. 1986) was implemented on Neuroscan 4.3 Edit software, in order to reduce the artefacts introduced by blinking and to re-reference data to the average reference. The average referenced data was used for further analysis. The data was subsequently epoched, with epochs encompassing 200ms before stimulus onset and up to 1500ms post-stimulus onset. The onset of the auditory stimulus marked the time 0 for baseline, peak analysis, and statistical purposes. For the VF and VS conditions time 0 was classified as the point where the auditory stimulus would have onset had it been presented. Further EEG analysis was undertaken with Brain Vision Analyzer 2.0 (Brain Products GmbH, München). The raw EEG data was digitally filtered with a low pass filter of 30Hz (slope 24dB/octave) to provide a bandwidth for analysis of 0.05–30Hz. The data was segmented according to condition and edited for possible sources of artefact using the following criteria: gradient criterion (maximum allowed voltage step: 50μV); amplitude criterion (±100μV); difference criterion (maximum allowed absolute difference: 200μV); and low activity criterion (lowest allowed activity max-min: 0.5μV, interval: 100ms). After artefact rejection, 99.8% of the original recordings were preserved. Epochs were baseline corrected to 100ms pre-auditory onset and averaged across the 60 trials separately for each modality (A, VF, AVF, VS, AVS). Only ERP responses to non-target trials were analysed.

Statistical analyses

Statistical analysis examined the effects of visual speech on auditory ERP responses and differences between TD and ASD groups. Significant AV interactions were examined by comparing responses in the auditory-only condition (A) to an estimation of the auditory response in the bimodal conditions once the contribution of the visual response had been removed, i.e. AVF minus VF (AVF-VF) and AVS minus VS (AVS-VS). Subtraction of the visual responses from the multisensory waveform is important as, in the absence of auditory

Autism Speaks Author Manuscript

Autism Speaks Author Manuscript

speech sounds, linguistic facial features are sufficient to activate auditory cortices (Calvert et al. 1997; MacSweeney et al. 2000). Significant differences between the auditory ERP responses in the A and AVF-VF, or A and AVS-VS, conditions were indicative of non-linear AV interactions (see Stein & Meredith, 1993). Topographical maps of the difference waves (AVF-VF minus A and AVS-VS minus A; see Figure 2) showed a pattern of AV effects consistent with activity within the superior temporal plane, i.e. positivity over central scalp and an inversion of polarity over the mastoids. AV interaction effects, with both visual speech and dynamic scrambled face stimuli, were observed over frontal and fronto-central scalp. However, over central and centro-parietal electrodes the AV effects appeared to be specific to auditory words accompanied by lip movements, particularly in the P2 time window (see Figure 2). On this basis two regions of interest were selected for statistical analysis. A fronto-central region (pooled electrodes F3, Fz, F4, FC1, FC2, FC3, FC4) and a centro-parietal region (pooled electrodes C3, Cz, C4, CP1, CP2, Pz). The ASD group also showed positivity over more anterior regions than the TD group, however, visual inspection of these waveforms revealed that auditory ERP peaks were not observed over these more frontal electrodes and effects were the result of slow wave AV differences. Therefore, the regions selected for analysis were based on the typical pattern of AV effects and confined to fronto-central and centro-parietal pooled electrodes. Auditory ERP responses were examined in three time windows based on grand average peaks: N1 (100–180ms post-auditory onset); P2 (200–300ms); and N4 (400–700ms). Data was entered into repeated measures ANOVAs with within-subject factors of region (2 levels: fronto-central, centro-parietal) and condition (3 levels: A, AVF-VF, AVS-VS). Differences between ASD and TD boys were examined by adding a between-subjects factor of group (2 levels: ASD, TD) and co-varying for PIQ to control for the significant difference in PIQ between TD and ASD groups.

The possibility that group differences in AV effects could be accounted for by low-level perceptual differences was explored. First, by examining group differences in unisensory auditory processing. Secondly, group differences in unisensory visual speech processing were examined by comparing ERP responses in the VF condition (silent lip movements) and the VS condition (visual-only scrambled face) over pooled electrodes in the N1, P2, and N4 time windows. Finally, potential group differences in N170 ERP responses (visual ERP response to the static face and static scrambled face preceding each AVF and AVS trial) were examined. Peak analysis was performed at electrode P8 (where these responses are maximal) in the time window 120ms to 250ms post-stimulus onset.

Results

Behavioural results

The hit rates and reaction times to detect the target trials were analysed using repeated measures ANOVAs with a within-subjects factor of condition (4 levels: A, AVF, AVS, VF), a between-subjects factor of Group (2 levels: ASD, TD), and PIQ as a covariate as TD boys had significantly higher PIQ than the ASD group. Accuracy rates were 97% across conditions and groups. For accuracy, there was no significant main effect of Condition ($p=0.4$), but there was a significant between-subjects effect of Group [$F(1, 25) = 5.6$, $p<0.05$]. However, in practice this amounted to a mean difference of one hit, with the ASD group scoring a mean of 23 hits (95% accuracy) relative to controls who detected a mean of 24 targets (99% accuracy). The TD group showed a mean false alarm rate of 1.4 over the whole course of the experiment, while for the ASD group the mean number of false alarms was 9.4. An ANOVA, with PIQ as a covariate, revealed no significant difference between the groups in the number of false alarms [$F(1, 25) = 2.5$, $p=0.1$].

Autism Speaks Author Manuscript

Autism Speaks Author Manuscript

For reaction times there was a significant main effect of Condition [F(3, 23) = 14.9, p<0.001] with faster reaction times in the AVF condition (805ms from start of naturally moving face) relative to the A condition (1507ms; p<0.001), AVS condition (964ms; p<0.001), or VF condition (1315ms; p<0.001). Reaction times were also significantly faster in the AVS condition relative to the unisensory conditions suggesting there is some behavioural advantage with even a task irrelevant AV stimuli. However, the significant difference between AVF and AVS conditions suggests an additional advantage was conferred by congruent lip movements. There was no significant between-subjects effect of Group (p=0.2). There was, however, a significant Condition x Group interaction [F(3, 23) = 3.2, p<0.05] but this appeared to be accounted for by slightly faster reaction times to the AVS targets for the ASD group (936ms) relative to the typical group (993ms).

Significant AV interactions and group differences

Analysis of the TD and ASD group together showed no significant AV interaction effects on the N1 component (F(2,24)=2.2, p=0.1) or the N4 component (F(2,24)=0.3, p=0.8). However, significant interaction effects were observed for the P2 component. These AV effects on the P2 were specific to the centro-parietal region (significant region by condition interaction: F(2,24)=7.2, p=0.004). When analysis was restricted to the centro-parietal region, a significant main effect of condition was observed (F(2,24)=4.2, p=0.03), with significantly greater P2 amplitude in the AVF-VF condition relative to the A condition (p=0.048). The ERP from the centro-parietal electrodes for the ASD group show clear differences to those obtained in the TD group (Figure 3). Most notable is the absence in the ASD group of the P2 amplitude enhancement in the AVF-VF condition that was observed in the TD group (significant condition by group interaction: F(2,24)=4.6, p=0.02). No significant condition by group interactions were found for N1 or N4 amplitude (p>0.2).

Significant AV interactions in TD adolescents

ERP waveforms (Figure 3) showed AV effects on all three ERP components (N1, P2, N4), however, the specificity of these AV effects to visual speech (the AVF-VF condition) differed between components, as did the topographical distribution of interaction effects.

An AV facilitation effect was found for N1 amplitude (F(2,12)=5.9, p=0.02) when data from fronto-central and centro-parietal regions was considered together, with significantly attenuated N1 amplitude in the AVF-VF condition relative to the A condition (p=0.004). However, the difference in N1 amplitude between AVF-VF and AVS-VS conditions was not significant (p=0.6).

There was also a significant main effect of condition on P2 amplitude (F(2,12)=10.0, p=0.003) with an amplitude enhancement in both AV conditions relative to the auditory-only condition (p<0.001 for AVF-VF and p=0.02 for AVS-VS). However, a significant region by condition interaction (F(2,12)=4.9, p=0.03) suggested topographical differences. Post-hoc tests on just the fronto-central region showed P2 amplitude enhancement (F(2,12)=12.8, p=0.001) in both AV conditions (p<0.001 for A and AVF-VF comparison and p=0.001 for A and AVS-VS). However, over the centro-parietal region the AV effect was specific to visual speech (F(2,12)=10.0, p=0.003) with significantly enhanced P2 amplitude in the AVF-VF condition relative to either the A (p=0.006) or the AVS-VS (p=0.03) condition.

Finally, an AV effect was found for N4 amplitude (F(2,12)=5.2, p=0.02) with significantly attenuated amplitude in the AVF-VF condition relative to the A condition (p=0.005). However, as for the N1 component, there was no significant difference in N4 amplitude between the AVF-VF and AVS-VS conditions (p=0.1).

Autism Speaks Author Manuscript

Autism Speaks Author Manuscript

Significant AV interactions in ASD adolescents

Analysis of the ASD group alone also revealed AV effects on all three ERP components. However, for the ASD group AV interactions in the P2 time window were driven by the scrambled face stimuli (the AVS-VS condition; see Figure 3).

At the level of the N1 the ASD group showed a trend for a significant main effect of condition ($F(2,12)=2.9$, $p=0.09$) with significantly attenuated N1 amplitude in the AVS-VS condition relative to the A condition ($p=0.03$). There was no significant difference in N1 amplitude between AVF-VF and AVS-VS conditions ($p=0.3$). There was also a significant region by condition interaction ($F(2,12)=5.6$, $p=0.02$) with only the fronto-central region showing a significant main effect of condition ($F(2,12)=5.6$, $p=0.02$) and significantly attenuated N1 amplitude in both AV conditions relative to the A condition ($p=0.03$ for AVF-VF and $p=0.005$ for AVS-VS).

For P2 amplitude a significant main effect of condition was found ($F(2,12)=8.9$, $p=0.004$) with significantly enhanced P2 amplitude in the AVS-VS condition relative to the A ($p=0.003$) or AVF-VF ($p=0.006$) conditions. There was also a significant region by condition interaction ($F(2,12)=4.4$, $p=0.04$). Post-hoc tests revealed no significant main effect of condition for the centro-parietal region ($F(2,12)=1.4$, $p=0.3$). Conversely, a significant main effect of condition was observed for the fronto-central region ($F(2,12)=17.1$, $p<0.001$) with significantly greater P2 amplitude in the AVS-VS condition relative to the A ($p<0.001$) or AVF-VF ($p=0.006$) condition. This P2 amplitude modulation when spoken words were accompanied by dynamic scrambled face stimuli was unexpected, and in contrast to the TD group.

There was also a significant region by condition interaction for N4 amplitude ($F(2,12)=4.8$, $p=0.03$) with only the fronto-central region showing AV effects in trend ($F(2,12)=3.9$, $p=0.051$) and significantly attenuated N4 amplitude in both AVF-VF ($p=0.04$) and AVS-VS ($p=0.01$) conditions relative to the A condition.

Group differences in AV interaction effects

The ERP from the centro-parietal electrodes for the ASD group show clear differences to those obtained in the TD group (Figure 3). Most notable is the absence in the ASD group of the P2 amplitude enhancement in the AVF-VF condition that was observed in the TD group (significant condition by group interaction: $F(2,24)=4.6$, $p=0.02$). No significant condition by group interactions were found for N1 or N4 amplitude ($p>0.2$).

Group differences in unisensory processing

To examine whether group differences in AV effects at the level of the P2 component could be accounted for by group differences in low-level unisensory processing we firstly examined ERP responses to the auditory-only spoken words. No significant main effect of group was found for auditory-only N1, P2, or N4 amplitude (all $p>0.3$). A significant region by group interaction ($F(1,25)=4.5$, $p=0.04$) was found for auditory-only P2 amplitude with a trend for the ASD group to show attenuated P2 amplitude in the A condition over the fronto-central region ($F(1,25)=3.9$, $p=0.06$). However, no group differences in P2 amplitude to the A condition were observed over the centro-parietal region, consequently the differences in AV effects cannot be accounted for by differences in auditory-only processing.

Secondly, group differences in unisensory visual processing (VF and VS conditions) were examined and no significant main effects of group, or group by condition interactions, were found in the N1, P2, or N4 time windows (all $p>0.4$).

Autism Speaks Author Manuscript

Autism Speaks Author Manuscript

Finally, we compared visual N170 ERP to the static face and static scrambled face which preceded each AVF and AVS trial. The TD group showed an N170 peak of −1.6μV (sd=2.3) at 175ms (sd=12.1) to the face, and of −0.9μV (sd=2.5) at 196ms (sd=31.7) for the scrambled face. The ASD group showed an N170 peak of −3.3μV (sd=3.9) at 195ms (sd=32.2) to the face, and of −0.4μV (sd=2.4) at 210ms (sd=40.3) to the scrambled face. There were no significant main effects of group for N170 latency (p=0.2) or amplitude (p=0.4).

Correlations between AV effects on P2 amplitude and clinical measures

The AV effects on P2 amplitude over the centro-parietal region were also compared against clinical measures (ADOS-G, Lord et al. 2000; SCQ, Rutter et al. 2003). When participants from both groups were included a significant negative correlation was found between SCQ score and AV effect (AVF-VF minus A) on P2 amplitude (Pearson's correlation=−0.5, p=0.003; see Figure 4). When analysis was restricted to the ASD group the negative correlation between SCQ score and AV effects on P2 amplitude remained significant (Pearson's correlation=−0.6, p=0.04). This suggests that the effects of visual speech on auditory P2 amplitude are reduced when autism symptomology is greater.

Discussion

We investigated audiovisual integration of whole spoken words in typically developing adolescents and high-functioning adolescents with ASD. As we will discuss below, we observed group differences in ERP AV effects, suggesting that the neural networks facilitating processing of spoken words by visual speech are altered in individuals with ASD. Moreover, the time window in which these group differences were observed was consistent with atypicalities in the lexical-semantic processing of AV words in ASD.

AV integration in typical development

A pattern of AV interaction effects consistent with activity within the superior temporal plane was found for TD adolescents, i.e. positivity over the central scalp and an inversion of polarity at the mastoids (see Figure 2). Interestingly, visual inspection of the topography of these AV interactions is consistent with two dissociable effects. For all time windows, there was a fronto-central AV effect showing general bimodal neural facilitation, both when spoken words were accompanied by a video of the speaker's face, and when participants heard words while viewing dynamic scrambled face stimuli. Conversely, the centro-parietal AV effect in the P2 time window (200–300ms post-auditory onset) was specific to viewing the speaker's face and lip movements. This P2 amplitude enhancement conflicts with previous reports of *decreased* P2 to AV speech (e.g. Klucharev et al. 2003; Stekelenburg & Vroomen, 2007; van Wassenhove et al. 2005). However, previous studies have used speech syllables so the use of word stimuli which have inherent meaning may account for the differential results, in particular as semantic expectations have been found to affect syllable processing in the same 200–300ms time window (Bonte et al. 2006). An important limitation of the current study is the absence of an extra control condition to remove spurious interaction effects (like anticipatory slow wave potentials), which will be present in the A condition, but removed from the AVF-VF condition. It is unlikely that the AV effects observed can be accounted for simply by reference to these effects as even if the A and AVF-VF difference could be accounted for, the differences between AVF-VF and AVS-VS remain significant, and indeed comparisons with no subtraction are highly significant. However, it is an important limitation and future studies would undoubtedly wish to add this extra control condition, such as the same fixation screen as in the A condition but without sound to produce a new equation: (A–C) - (AVF-VF) as used by Stekelenburg and Vroomen (2007).

Autism Speaks Author Manuscript

Autism Speaks Author Manuscript

ASD differences in AV integration

The ASD group showed AV effects over the fronto-central scalp in the N1 and N4 time windows with attenuated amplitude responses to both bimodal conditions as in the TD group. Unexpectedly, however, the ASD individuals showed a significant modulation of fronto-central P2 amplitude by socially uninformative stimuli (scrambled faces). Interesting, a significant group difference was also observed in reaction times to the scrambled face targets with the ASD group showing significantly shorter reaction times to the AVS stimuli than the TD group. Moreover, group differences were found in the modulation of P2 amplitude over the centro-parietal region. Crucially, this is the region and the AV effect which showed specificity to visual speech (over and above dynamic accompanying visual information) in the TD group. However, it is important to note that although a target detection task was used in order to draw attention to the stimuli, there were no explicit fixation instructions, and thus gaze differences between groups could have led to differential attention and hence account for the group difference. Importantly, however, group differences were not found in ERP responses to the unisensory visual-only (or auditory-only) speech conditions suggesting that reduced AV effects on P2 amplitude in ASD are specific to AV *integration*, and cannot be accounted for on the basis of lack of attention to the faces, or deficits in lip reading as found by Iarocci et al. (2010).

The functional significance of the P2 ERP component is less well understood than the N1 component which has been associated with sound onset detection (see Čeponiené et al. 2005; Sanders et al. 2002) or the N4 component where amplitude modulation indexes semantic integration (Domalski et al. 1991; Kutas & Hillyard, 1980; Rugg, 1985). As Pulvermüller et al. (2009) discuss, this may be because early (<250ms) semantic effects on ERP responses are smaller, shorter-lasting and more focal than effects on the N4 component. The P2 component has, however, been shown to be sensitive to semantic priming. For instance, Bentin et al. (1985) found enhanced P2 amplitude to visual words which were preceded by a prime that was a word from the same category (e.g. rain-snow). Thus the group differences found in this study suggest that AV integration during the lexical-semantic processing of speech is atypical in ASD. This is consistent with behavioural findings in high-functioning ASD individuals that semantic deficits persist into adolescence and adulthood, particularly for the comprehension of auditory verbal information (Paul & Cohen, 1985; Strandburg et al. 1993; Tager-Flusberg, 1991). Interestingly, the magnitude of the AV effects on P2 amplitude was found to negatively correlate with scores on the Social Communication Questionnaire (SCQ; Rutter et al. 2003), suggesting that reduced neural facilitation in a lexical-semantic processing window was associated with increased autism symptomology. Magnée et al. (2008) failed to find ASD group differences in AV effects on P2 amplitude in response to spoken syllables. However, these conflicting results could be explained by the different stimulus material used, i.e. the effects of semantic expectations (discussed previously) or the different electrodes examined. The finding of group differences in the lexical-semantic processing of AV speech suggests that the ASD group are not using the lip movements to facilitate in the processing of auditory speech. The clinical implications of this difference are that audiovisual speech may be processed in a less ‘integrated’ way than is typical. Thus to use an analogy the ASD group may ‘see what you are saying’ but not necessarily ‘see what you mean’ in that lip movements may be used to predict the timing of the onset of auditory speech (as reflected by no significant group differences at the level of the N1) but when it comes to processing for meaning the information conveyed by the lip movements is not being used to facilitate the lexical-semantic processing of the auditory word. The reasons for the different use of visual speech remain to be explored but might include different audiovisual processing strategies, e.g. adopting a ‘wait-and-see’ strategy when it comes to processing auditory words for meaning rather than using information from the lip movements to predict the identity of the auditory

Autism Speaks Author Manuscript

Autism Speaks Author Manuscript

word, or differences in underlying brain structures responsible for AV processing. The topography of AV effects on the P2 in TD adolescents is consistent with activity within the superior temporal plane. Imaging studies have implicated the superior temporal sulcus (STS) region in the perception of biological motion (e.g. Howard et al. 1996; Bonda et al. 1996), voice perception (e.g. Belin et al. 2000; Belin & Zatorre, 2003), and complex social cognition (e.g. Castelli et al. 2000; Schultz et al. 2004). Interestingly, no significant AV effects were observed in the ASD group over the centro-parietal region, and this may be consistent with findings of anatomical and functional STS abnormalities in ASD (see Zilbovicius et al. 2006, for review).

Acknowledgments

Grant Sponsor: MRC, Autism Speaks Grant Number: 1280

This study was supported by the Medical Research Council and Autism Speaks. We wish to thank all the participants and their parents, as well as Deiniol Buxton, Emma and Mike Askem, and Sundeep Dhesi for their role in the creation of the stimulus battery.

References

Baird G, Simonoff E, Pickles A, Chandler S, Loucas T, Meldrum D, Charman T. Prevalence of disorders of the autism spectrum in a population cohort of children south thames: the special needs and autism project (SNAP). The Lancet. 2006; 368:210–215.

Bebko JM, Weiss JA, Demark JL, Gomez P. Discrimination of temporal synchrony in intermodal events by children with autism and children with developmental disabilities without autism. Journal of Child Psychology and Psychiatry. 2006; 47:88–98. [PubMed: 16405645]

Belin P, Zatorre RJ. Adaptation to speaker's voice in right anterior temporal lobe. Neuro Report. 2003; 14:2105–2109.

Belin P, Zatorre RJ, Lafaille P, Ahad P, Pike B. Voice-selective areas in human auditory cortex. Nature. 2000; 403:309–312. [PubMed: 10659849]

Bentin S, McCarthy G, Wood CC. Event-related potentials, lexical decision and semantic priming. Electroencephalography and Clinical Neurophysiology. 1985; 60:343–355. [PubMed: 2579801]

Besle J, Fort A, Delpuech C, Giard MH. Bimodal speech: early suppressive visual effects in human auditory cortex. European Journal of Neuroscience. 2004; 20:2225–2234. [PubMed: 15450102]

Bonda E, Petrides M, Ostry D, Evans A. Specific involvement of human parietal systems and the amygdale in the perception of biological motion. Journal of Neuroscience. 1996; 16:3737–3744. [PubMed: 8642416]

Bonte M, Parviainen T, Hytönen K, Salmelin R. Time course of top- down and bottom-up influences on syllable processing in the auditory cortex. Cerebral Cortex. 2006; 16:115–123. [PubMed: 15829731]

Brown GDA. A frequency count of 190,000 words in the london–lund corpus of english conversation. Behavior Research Methods, Instruments, & Computers. 1984; 16:502–532.

Calvert GA, Bullmore ET, Brammer MJ, Campbell R, Williams SCR, McGuire PK, Woodruff PWR, Iversen SD, David AS. Activation of auditory cortex during silent lipreading. Science. 1997; 276:593–596. [PubMed: 9110978]

Castelli F, Happé F, Frith U, Frith C. Movement and mind: a functional imaging study of perception and interpretation of complex intentional movement patterns. Neuro Image. 2000; 12:314–325. [PubMed: 10944414]

Čeponiené R, Alku P, Westerfield M, Torki M, Townsend J. ERPs differentiate syllables and nonphonetic sound processing in children and adults. Psychophysiology. 2005; 42:391–406. [PubMed: 16008768]

de Gelder B, Vroomen J, Van der Heide L. Face recognition and lip- reading in autism. European Journal of Cognitive Psychology. 1991; 3:69–86.

Autism Speaks Author Manuscript

Autism Speaks Author Manuscript

Domalski P, Smith ME, Halgren E. Cross-modal repetition effects on the N4. Psychological Science. 1991; 2:173–178.

Dunn M, Bates J. Developmental change in neural processing of words by children with autism. Journal of Autism Developmental Disorders. 2005; 35(3):361–376.

Dunn M, Vaughan H, Kreuzer J, Kurtzberg D. Electrophysiological correlates of semantic classification in autistic and normal children. Developmental Neuropsychology. 1999; 16:79–99.

Frith U, Snowling M. Reading for meaning and reading for sound in autistic and dyslexic children. British Journal of Developmental Psychology. 1983; 1:329–42.

Happé FGE. Central coherence and theory of mind in autism: Reading homographs in context. British Journal of Developmental Psychology. 1997; 15:1–12.

Harris GJ, Chabris CF, Clark J, Urban T, Aharon I, Steele S, McGrath L, Condouris K, Tager-Flusberg H. Brain activation during semantic processing in autism spectrum disorders via functional magnetic resonance imaging. Brain and Cognition. 2006; 61:54–68. [PubMed: 16473449]

Howard RJ, Brammer M, Wright I, Woodruff PW, Bullmore ET, Zeki S. A direct demonstration of functional specialization within motion-related visual and auditory cortex of the human brain. Current Biology. 1996; 6:1015–1019. [PubMed: 8805334]

Iarocci G, Rombough A, Yager J, Weeks DJ, Chua R. Visual influences on speech perception in children with autism. Autism. 2010; 14:305–320. [PubMed: 20591957]

Klucharev V, Möttönen R, Sams M. Electrophysiological indicators of phonetic and non-phonetic multisensory interactions during audiovisual speech perception. Cognitive Brain Research. 2003; 18:65–75. [PubMed: 14659498]

Kutas M, Federmeier KD. Electrophysiology reveals semantic memory use in language comprehension. Trends in Cognitive Sciences. 2000; 4:463–470. [PubMed: 11115760]

Kutas M, Hillyard SA. Reading senseless sentences: brain potentials reflect semantic incongruity. Science. 1980; 207:203–205. [PubMed: 7350657]

Lord C, Risi S, Lambrecht L, Cook EH Jr, Leventhal BL, DiLavore PC, Pickles A, Rutter M. The autism diagnosis observation schedule – generic: a standard measure of social and communication deficits associated with the spectrum of autism. Journal of Autism and Developmental Disorders. 2000; 30:205–223. [PubMed: 11055457]

Lord C, Rutter M, Le Couteur A. 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. 1994; 24:659–685. [PubMed: 7814313]

McGurk H, MacDonald J. Hearing lips and seeing voices. Nature. 1976; 264:746–748. [PubMed: 1012311]

McPartland J, Dawson G, Webb SJ, Panagiotides H, Carver LJ. Event-related brain potentials reveal anomalies in temporal processing of faces in autism spectrum disorder. Journal of Child Psychology and Psychiatry. 2004; 45:1235–1245. [PubMed: 15335344]

MacSweeney M, Amaro E, Calvert GA, Campbell R, David AS, McGuire P, Williams SCR, Woll B, Brammer MJ. Silent speechreading in the absence of scanner noise: an event-related fMRI study. Neuro Report. 2000; 11:1729–1733.

Magnée MJCM, de Gelder B, van Engeland H, Kemner C. Audiovisual speech integration in pervasive developmental disorder: evidence from event-related potentials. Journal of Child Psychology and Psychiatry. 2008; 49:995–1000. [PubMed: 18492039]

Norbury CF. The relationship between theory of mind and metaphor: evidence from children with language impairment and autistic spectrum disorder. British Journal of Developmental Psychology. 2005; 23:383–399.

Osterling J, Dawson G. Early recognition of children with autism: a study of first birthday home videotapes. Journal of Autism and Developmental Disorders. 1994; 24:247–257. [PubMed: 8050980]

Paul R, Cohen D. Comprehension of indirect requests in adults with autistic disorders and mental retardation. Journal of Speech and Hearing Research. 1985; 28:475–479. [PubMed: 4087881]

Pulvermüller F, Shtyrov Y, Hauk O. Understanding in an instant: Neurophysiological evidence for mechanistic language circuits in the brain. Brain and Language. 2009; 110:81–94. [PubMed: 19664815]

Autism Speaks Author Manuscript

Autism Speaks Author Manuscript

Ring H, Sharma S, Wheelwright S, Barrett G. An electrophysiological investigation of semantic incongruity processing by people with asperger's syndrome. Journal of Autism and Developmental Disorders. 2007; 37:281–290. [PubMed: 16865545]

Rugg M. The effects of semantic priming and word repetition on event- related potentials. Psychophysiology. 1985; 22:642–647. [PubMed: 4089090]

Rutter, M.; Bailey, A.; Berument, SK.; Le Couteur, A.; Lord, C.; Pickles, A. Social Communication Questionnaire (SCQ). Los Angeles, CA: Western Psychological Services; 2003.

Sanders LD, Newport EL, Neville HJ. Segmenting nonsense: an event- related potential index of perceived onsets in continuous speech. Nature Neuroscience. 2002; 5:700–703.

Schultz J, Imamizu H, Kawato M, Frith CD. Activation of the human superior temporal gyrus during observation of goal attribution by intentional objects. Journal of Cognitive Neuroscience. 2004; 10:1695–1705. [PubMed: 15701222]

Semlitsch HV, Anderer P, Schuster P, Presslich O. A solution for reliable and valid reduction of ocular artifacts, applied to the P300 ERP. Psychophysiology. 1986; 23:695–703. [PubMed: 3823345]

Smith EG, Bennetto L. Audiovisual speech integration and lipreading in autism. Journal of Child Psychology and Psychiatry. 2007; 48:813–821. [PubMed: 17683453]

Snowling M, Frith U. Comprehension in 'hyperlexic' readers. Journal of Experimental Child Psychology. 1986; 42:392–415. [PubMed: 3806010]

Stein, BE.; Meredith, MA. The Merging of the Senses. London: MIT Press; 1993.

Stekelenburg JJ, Vroomen J. Neural correlates of multisensory integration of ecologically valid audiovisual events. Journal of Cognitive Neuroscience. 2007; 19:1964–1973. [PubMed: 17892381]

Strandburg RJ, Marsh JT, Brown WS, Asamow RF, Guthrie D, Higa J. Event-related potentials in high-functioning adult autistics: linguistic and nonlinguistic visual information processing tasks. Neuropsychologia. 1993; 31:413–434. [PubMed: 8502377]

Tager-Flusberg H. Semantic processing in the free recall of autistic children: further evidence for a cognitive deficit. British Journal of Developmental Psychology. 1991; 9:417–430.

van Wassenhove V, Grant KW, Poeppel D. Visual speech speeds up the neural processing of auditory speech. Proceedings of the National Academy of Sciences United States of America. 2005; 102:1181–1186.

Wechsler, D. WASI Manual. San Antonio, TX: The Psychological Corporation; 1999.

Williams JH, Massaro DW, Peel NJ, Bosseler A, Suddendorf T. Visual auditory integration during speech imitation in autism. Research in Developmental Disabilities. 2004; 25:559–575. [PubMed: 15541632]

Zilbovicius M, Meresse I, Chabane N, Brunelle F, Samson Y, Boddaert N. Autism, the superior temporal sulcus and social perception. Trends in Neurosciences. 2006; 29:359–366. [PubMed: 16806505]

Autism Speaks Author Manuscript

Autism Speaks Author Manuscript

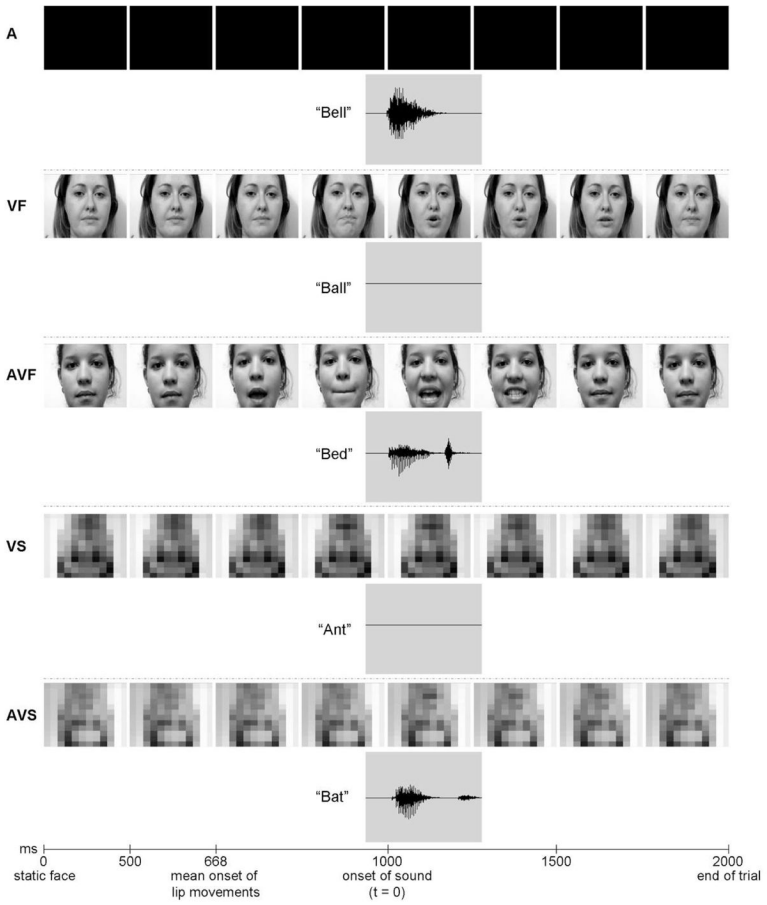

Figure 1.
Examples of stimuli used in the different experimental conditions: spoken words in A (auditory-only), VF (visual-only with face), AVF (audiovisual with face), VS (visual-only with scrambled face), and AVS (audiovisual with scrambled face) conditions.

Autism Speaks Author Manuscript

Autism Speaks Author Manuscript

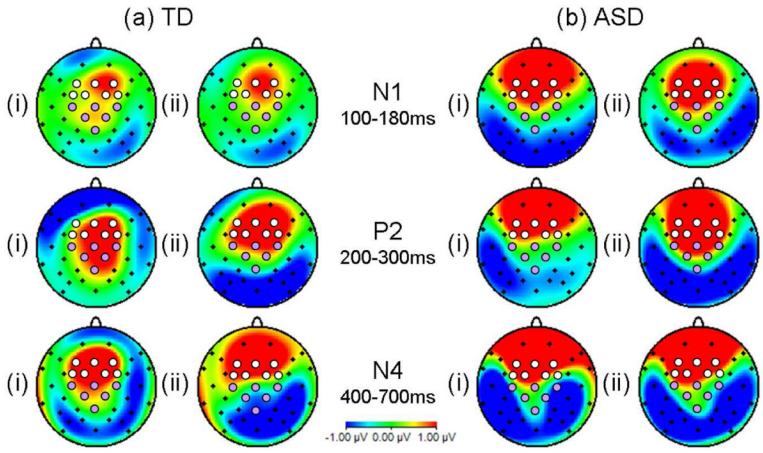

Figure 2.

Maps indicating topography of AV effects, i.e. difference waves for (i) AVF-VF minus A and (ii) AVS-VS minus A in N1 (100–180ms), P2 (200–300ms), and N4 (400–700ms) time windows. Electrodes selected for ERP data analysis are indicated by white (fronto-central region) and mauve (centro-parietal) circles.

Autism Speaks Author Manuscript

Autism Speaks Author Manuscript

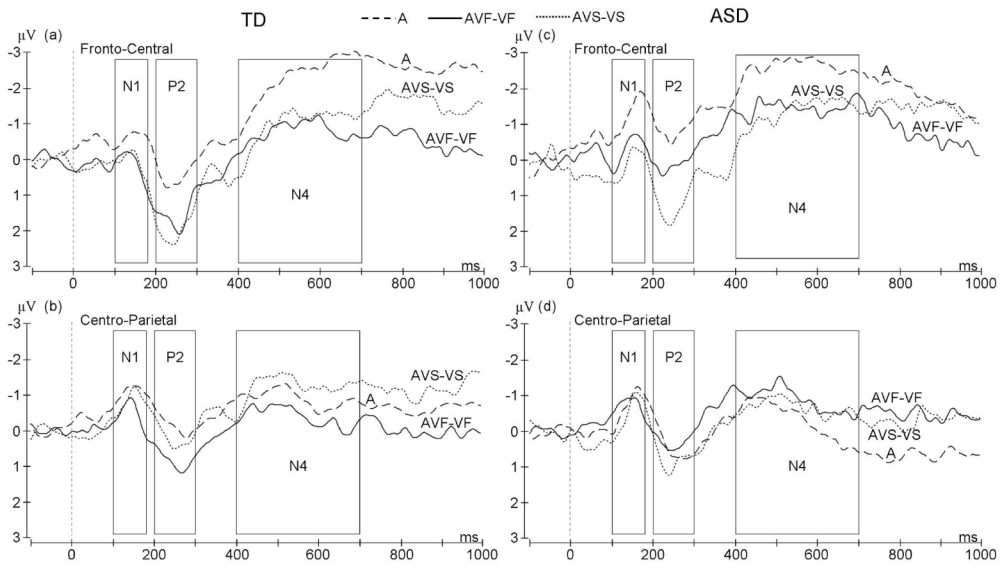

Figure 3.

Grand average ERP waveform showing AV effects on the N1, P2, and N4 components for the TD group at (a) pooled fronto-central electrodes and (b) pooled centro-parietal electrodes and for the ASD group at (c) fronto-central and (d) centro-parietal electrodes.

Autism Speaks Author Manuscript

Autism Speaks Author Manuscript

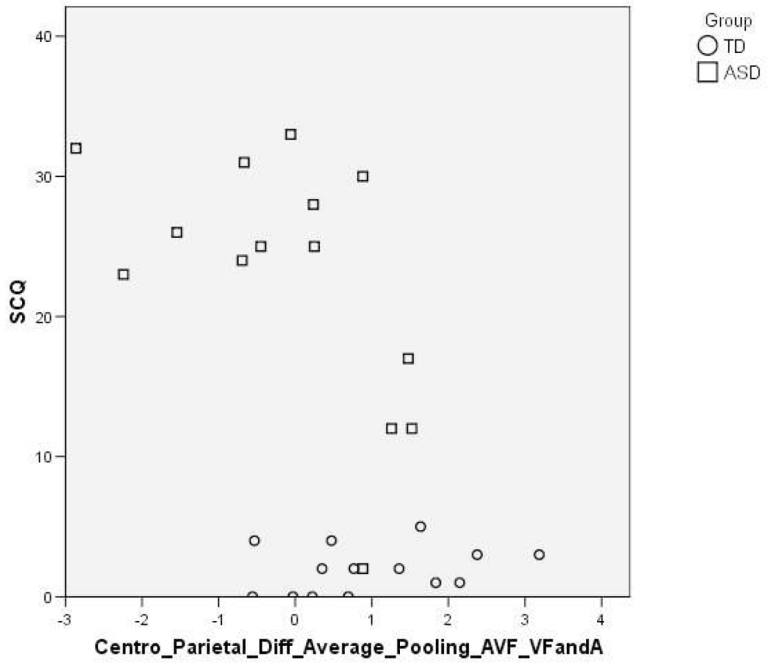

Figure 4.
Scatterplot showing the significant correlation between AV effects on P2 amplitude over centro-parietal electrodes (difference between AVF-VF and A) and SCQ score (with the TD group represented by circles and the ASD group as squares).

Autism Speaks Author Manuscript

Autism Speaks Author Manuscript

Table 1

	ASD GROUP (N=14)	TD GROUP (N=14)
AGE	16.9 (0.3)	16.9 (0.9)
VIQ	96.1 (8.4)	101.5 (10.1)
PIQ	103.4 (8.6)	109.8 (6.4)