NeuroImage: Clinical 20 (2018) 1078–1091

Contents lists available at ScienceDirect

NeuroImage: Clinical

journal homepage: www.elsevier.com/locate/ynicl

Recognizing visual speech: Reduced responses in visual-movement regions, but not other speech regions in autism

Kamila Borowiak[a,b,c,*], Stefanie Schelinski[a,c], Katharina von Kriegstein[a,c]

[a] Max Planck Institute for Human Cognitive and Brain Sciences, Stephanstraße 1a, 04103 Leipzig, Germany
[b] Berlin School of Mind and Brain, Humboldt University of Berlin, Luisenstraße 56, 10117 Berlin, Germany
[c] Technische Universität Dresden, Bamberger Straße 7, 01187 Dresden, Germany

ARTICLE INFO

Keywords:
High-functioning autism
Lip reading
Atypical perception
Motion
fMRI
Face

ABSTRACT

Speech information inherent in face movements is important for understanding what is said in face-to-face communication. Individuals with autism spectrum disorders (ASD) have difficulties in extracting speech information from face movements, a process called visual-speech recognition. Currently, it is unknown what dysfunctional brain regions or networks underlie the visual-speech recognition deficit in ASD.

We conducted a functional magnetic resonance imaging (fMRI) study with concurrent eye tracking to investigate visual-speech recognition in adults diagnosed with high-functioning autism and pairwise matched typically developed controls.

Compared to the control group ($n = 17$), the ASD group ($n = 17$) showed decreased Blood Oxygenation Level Dependent (BOLD) response during visual-speech recognition in the right visual area 5 (V5/MT) and left temporal visual speech area (TVSA) – brain regions implicated in visual-movement perception. The right V5/MT showed positive correlation with visual-speech task performance in the ASD group, but not in the control group. Psychophysiological interaction analysis (PPI) revealed that functional connectivity between the left TVSA and the bilateral V5/MT and between the right V5/MT and the left IFG was lower in the ASD than in the control group. In contrast, responses in other speech-motor regions and their connectivity were on the neurotypical level.

Reduced responses and network connectivity of the visual-movement regions in conjunction with intact speech-related mechanisms indicate that perceptual mechanisms might be at the core of the visual-speech recognition deficit in ASD. Communication deficits in ASD might at least partly stem from atypical sensory processing and not higher-order cognitive processing of socially relevant information.

1. Introduction

In face-to-face communication, fast and accurate perception of the visible articulatory movements in the face can substantially enhance our understanding of auditory speech (Sumby and Pollack, 1954; Van Wassenhove et al., 2005). This is particularly beneficial for perceiving speech in noisy environments (Ross et al., 2007), and for hearing-impaired populations (Giraud et al., 2001; Rouger et al., 2007).

Difficulties in recognizing visual speech likely contribute to communication difficulties that are one of the core symptoms in autism spectrum disorders (ASD, DSM-5, American Psychiatric Association, 2013). To date, there is a large body of behavioral evidence for visual-speech recognition deficits in ASD (e.g., Williams et al., 2004; Smith and Bennetto, 2007; Schelinski et al., 2014), but brain regions or networks that might underlie the behavioral difficulties remain unknown.

Visual-speech recognition in the typically developed population involves several brain regions (Calvert et al., 1997; Campbell et al., 2001; Okada and Hickok, 2009; Blank and von Kriegstein, 2013). These regions can be broadly divided into “visual-movement regions” for the processing of visual movement, and “speech-motor regions” involved in production and perception of auditory speech (Wilson et al., 2004; Skipper et al., 2005). Visual-movement regions refer to the motion-sensitive areas in the V5/MT and the posterior superior temporal sulcus/gyrus (pSTS/STG). V5/MT is an extrastriate visual area sensitive to human and non-human movement (Zeki et al., 1991; Beckers and Homberg, 1992; Grèzes et al., 2001). The pSTS/STG is relevant for human motion perception (Puce et al., 1998; Grossman et al., 2005), and the left pSTS/STG particularly for processing visual speech (Hall et al., 2005; Lee et al., 2007). The visual-speech sensitive portion of the pSTS/STG has been coined the temporal visual speech area (TVSA;

* Corresponding author at: Technische Universität Dresden, Bamberger Straße 7, 01187 Dresden, Germany.
E-mail address: borowiak@cbs.mpg.de (K. Borowiak).

https://doi.org/10.1016/j.nicl.2018.09.019
Received 9 May 2018; Received in revised form 19 September 2018; Accepted 21 September 2018
Available online 24 September 2018

Bernstein et al., 2011). Speech-motor regions include the left inferior frontal gyrus (IFG), the precentral gyrus (PCG) and the supplementary motor area (SMA). Activation of speech-motor regions during visual-speech perception is subsequent to activation in visual-movement regions (Nishitani and Hari, 2002; Chu et al., 2013). The two groups of brain regions might reflect two stages of visual-speech recognition: perception of motion signals in the face ("visual-movement regions"), and a subsequent stage of extracting the speech information from the motion ("speech-motor regions").

In light of the current debate about the role of sensory processing for communication deficits in ASD (Baum et al., 2015; Robertson and Baron-Cohen, 2017), our study aim was to find out, whether visual-speech recognition difficulties in ASD are due to atypical brain mechanisms for the perception of human motion (Blake et al., 2003; Herrington et al., 2007), or rather subsequent mechanisms for speech processing (Boddaert et al., 2003; Tryfon et al., 2018). If the first option is true, we expect lower responses and/or lower connectivity in the visual-movement regions in ASD in contrast to controls during visual-speech recognition. In contrast, if the latter is true, we expect neurotypical responses and connectivity in visual-movement regions, but lower responses and/or connectivity in the speech-motor regions in ASD in contrast to controls.

We used functional magnetic resonance imaging (fMRI) and concurrent eye tracking to systematically investigate visual-speech recognition in adults with high-functioning ASD and typically developed pairwise matched controls. In an fMRI visual-speech recognition experiment, participants saw silent videos of speakers articulating syllables and performed a visual–speech task and a face-identity task. The two tasks were performed on identical stimulus material. Contrasting the visual-speech task to the face-identity task allowed us to specifically target mechanisms underlying visual-speech processing in contrast to processing of other face information. We applied an fMRI region of interest (ROI) localizer to functionally localize the motion-sensitive V5/MT and the TVSA in the left pSTS/STG (von Kriegstein et al., 2008). Tracking participants' eye movements during visual-speech recognition in the MRI environment was motivated by previous studies, which reported that individuals with ASD gaze less to the face and the mouth during visual-speech recognition compared to typically developed controls (Irwin et al., 2011; Irwin and Brancazio, 2014, but see Foxe et al., 2015). Gaze behavior is an important factor to consider because eye movements to informative parts of the face are a prerequisite for successful visual-speech recognition (Marassa and Lansing, 1995), and influence brain responses to visual-speech (Jiang et al., 2017).

2. Materials and methods

2.1. Participants and neuropsychological assessment

The study sample included 17 individuals diagnosed with ASD (ASD group) and 17 typically developed individuals (control group) who were matched pairwise on gender, chronological age, handedness (Oldfield, 1971) and full performance intelligence quotient (IQ) (Table 1). We excluded three additional participants with ASD: one participant due to difficulties in finding a control subject who would match with regard to IQ (full scale IQ = 85), one participant due to head movements in the MRI scanner greater than 3 mm during the visual-speech recognition experiment, and one participant due to a performance in the visual-speech recognition experiment that was lower than 2 standard deviations of the mean performance of the ASD group. Data of the respective control participants was excluded as well.

All participants were on a high-functioning cognitive level as indicated by an IQ within the normal range or above (defined as a full scale IQ of at least 85). Pairs of ASD and control participants were considered matched on IQ if the full scale IQ difference within each pair was maximally one standard deviation (15 IQ points). IQ was assessed using the Wechsler Adult Intelligence Scale (WAIS III; Wechsler, 1997; German adapted version: von Aster et al., 2006). In addition, groups showed comparable concentration performances (d2 test of attention; Brickenkamp, 2002; Table 1). All participants reported normal or corrected-to-normal vision correction. All reported normal hearing abilities and we confirmed these reports by means of pure tone audiometry (hearing level equal or below 35 dB at the frequencies of 250, 500, 1000, 1500, 2000, 3000, 4000, 6000, and 8000 Hz) (Micromate 304; Madsen, Denmark). All participants were native German speakers and were free of psychostimulant medication.

Table 1
Descriptive statistics for the ASD (n = 17) and the control group (n = 17) and group comparisons. Each participant in the control group was matched with respect to chronological age, gender, intelligence quotient (IQ), and handedness to the profile of one ASD participant.

	Control (n = 17)		ASD (n = 17)		
Gender	13 males, 4 females		13 males, 4 females		
Handedness[a]	14 right, 3 left		14 right, 3 left		
	M	*SD (range)*	*M*	*SD (range)*	*p*
Age	32.65	11.08 (21–55)	31.47	10.82 (21–54)	0.756
WAIS-III[b] scales					
Full scale IQ	107.12	8.17 (91–121)	105.35	10.64 (87–124)	0.591
Verbal IQ	106.29	10.84 (89–130)	109.06	12.61 (91–138)	0.498
Performance IQ	106.76	8.78 (90–121)	100.12	9.76 (82–120)	0.045*
Working memory	103.76	11.44 (88–126)	105.65	13.32 (86–146)	0.662
Concentration[c]	105.12	7.66 (86–114)	101.82	11.73 (84–126)	0.341
AQ[d]	17.06	4.07 (10–25)	37.94	7.82 (14–47)	0.000*

[a] Handedness was assessed using the Edinburgh handedness questionnaire (Oldfield, 1971).
[b] WAIS-III = German adapted version of the Wechsler Adult Intelligence Scale (Wechsler, 1997; M = 100; SD = 15).
[c] Concentration = d2 test of attention (Brickenkamp, 2002; M = 100; SD = 10).
[d] AQ = Autism Spectrum Quotient (Baron-Cohen et al., 2001).
* Significant group differences ($p < .05$); M = mean; SD = standard deviation.

Participants with ASD had previously received a formal clinical diagnosis of Asperger Syndrome (13 male, 4 female) or childhood autism (1 male, verbal IQ 119) according to the diagnostic criteria of the International Classification of Diseases (ICD; World Health Organization, 2004). The diagnosis was additionally confirmed based on the Autism Diagnostic Observation Schedule (ADOS; Lord et al., 2000; German version: Rühl et al., 2004), that was conducted in the context of clinical diagnostics or by trained researchers (KB, SS). If caregivers or relatives were available (n = 11), we also performed the Autism Diagnostic Interview-Revised (ADI-R; Lord et al., 1994; German version: Bölte et al., 2003). Five ASD participants had previously received a formal clinical diagnosis of other comorbid psychiatric disorders (social anxiety, depression (remitted) and posttraumatic stress disorder) according to the diagnostic criteria of the ICD (World Health Organization, 2004). Control participants were screened for presence of autistic traits and none of them met a clinically relevant extend as assessed by the Autism Spectrum Quotient (AQ; Baron-Cohen et al., 2001, Table 1). Note that one control participant had a higher AQ score than one of the ASD participants. This is expected since the distribution of the AQ score has been shown to overlap between the ASD and the neurotypical population (Baron-Cohen et al., 2001). The AQ is a self-assessment screening instrument for measuring the degree of autistic traits, but it does not serve as a diagnostic tool. It is suitable to

discriminate between individuals diagnosed with ASD and neurotypical controls (e.g., Baron-Cohen et al., 2001; Wakabayashi et al., 2006), but it does not significantly predict a receipt of ASD diagnosis (Ashwood et al., 2016). None of the control participants reported any history of psychiatric disorders or any family history of ASD. None of the participants reported any history of neurological disease. Written informed consent was obtained from all participants according to the procedures approved by the Ethics Committee of the Medical Faculty at the University Leipzig (316-15-24082015). All participants received expense reimbursement (8€/hour for MRI session, 7€/hour for behavioral session and travel cost reimbursement).

2.2. Experiments

2.2.1. Visual-speech recognition experiment (fMRI)

The experiment was a 2 × 2 factorial design with the factors *Task* (visual-speech task, face-identity task) and *Group* (control, ASD). The stimulus material consisted of silent videos of speakers articulating a vowel-consonant-vowel (VCV) syllable. The videos were taken from 3 male speakers and there were 63 different syllables for each speaker. The syllables represented all combinations of the consonants /f/, /l/, /n/, /p/, /r/, /s/, /t/ and the vowels /a/, /e/, /u/. Syllables were pseudorandomly assorted into blocks of nine videos considering the German viseme classes (Aschenberner and Weiss, 2005). In each block, the participants either performed the visual-speech task or the face-identity task (Fig. 1A). Before each block (Fig. 1B), participants received a task instruction: They saw a written instruction screen "syllable" or "person" to announce which task to perform. The screen was followed by the presentation of one video of one of the 3 speakers articulating one of the syllables. For the visual-speech task, participants were asked to memorize the syllable of this video (target syllable) and to indicate for each of the videos in the block whether the syllable matched the target syllable or not, independent of the person who was articulating it. For the face-identity task, participants were asked to memorize the person in the video (target person) and to decide for each video within the block, whether the person matched the target person or not, independent of the syllable that was articulated. After each block, a white fixation cross on a black screen was presented for a period of 18 s. The stimulus material was exactly the same for both tasks. There were 21 blocks in the visual-speech task and 21 blocks in the face-identity task. Blocks and trials within a block were presented in a pseudorandomized order. The number of target items varied between two and five across blocks and was the same for the visual-speech task and the face-identity task. Responses were made via a button box. Participants were requested to respond to each item by pressing one button if it was a target and another button if it was not. The experiment was divided into two fMRI runs of 15 min.

Before the fMRI experiment, participants were familiarized with the visual-speech task and the face-identity task outside the MRI scanner. They conducted 3 practice blocks per task, which had the same structure as blocks for the actual scanning session, but a different stimulus material (3 speakers and 9 VCV-syllables not included in the fMRI experiment).

All videos started and ended with a closed mouth of the speaker providing all movements made during syllable production. Videos were on average 2.18 s (± 0.12 s) long. Syllables were recorded from six professional male native German speakers who were all unfamiliar to the participants (24, 25, 26, 26, 27 and 31 years old). Three speakers were presented in the test phase and the other three speakers were used for the purpose of task familiarization. All speakers articulated the same set of syllables in a neutral manner and under the same conditions. Only the head of the speakers was displayed face-on against a uniform black background. Videos were recorded with a digital video camera (Canon-Legria HFS100, Canon Inc., Tokyo, Japan) and edited in Final Cut Pro (version 7, Apple Inc., CA, USA). Videos were overlaid with a mask so that outer features of the face (i.e. hair and ears) and the background were blurred. Videos were converted to grayscale and AVI 4:3 format (1024 × 768 pixels).

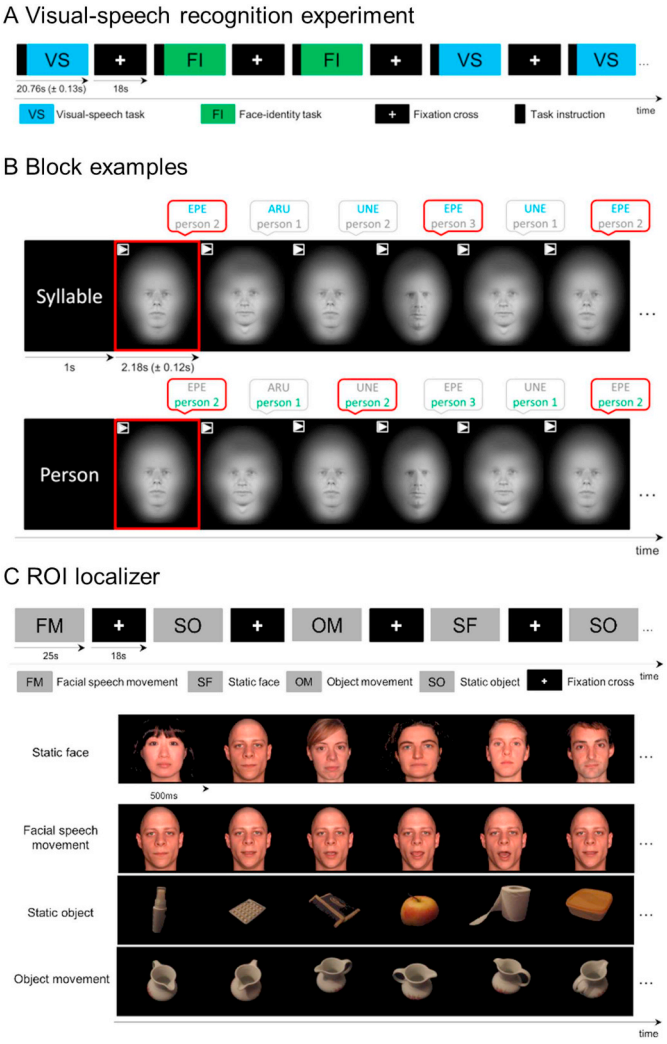

Fig. 1. Experimental designs presented during fMRI session. **(A)** Visual-speech recognition experiment: Participants viewed blocks of videos without audiostream showing 3 speakers articulating syllables. There were two tasks for which the same stimuli were used: visual-speech task and face-identity task. **(B)** At the beginning of each block, a written word instructed participants to perform one of the tasks ("syllable" for the visual-speech task or "person" for the face-identity task). In the visual-speech task, participants matched the articulated syllable to a target syllable (here 'EPE'). In the face-identity task, participants matched the identity of the speaker to a target person (here person 2). Respective targets were presented in the first video of the block and marked by a red frame around the video. **(C)** ROI localizer: Blocks of images of faces and objects were presented, and participants were asked to view them attentively.

2.2.2. ROI localizer (fMRI)

The ROI localizer was a 2 × 2 × 2 factorial design with the factors *Stimulus* (face, object), *Movement* (static, movement) and *Group* (control, ASD). It was based on the design by von Kriegstein et al. (2008). The localizer included four conditions (Fig. 1C): (i) static face (faces from different persons with different facial gestures while speaking), (ii) facial speech movement (different facial gestures of the same person's face while speaking), (iii) static object (different objects in different views), and (iv) object movement (same object in different views). In conditions (i) and (iii) the stream of pictures gave the impression of individual faces or objects, while conditions (ii) and (iv) induced the impression of one speaking face or moving object. Participants were asked to attentively view blocks of pictures of faces and

objects. Each block lasted 25 s and within the blocks, each single picture was presented for 500 ms without any pause between stimuli. After each block, a white fixation cross on a black screen was presented for a period of 18 s. There were four blocks per condition presented in two fMRI runs of 6 min.

2.2.3. Behavioral tests

To assess visual-speech recognition independent of the fMRI experiment, participants performed a visual word-matching test (Schelinski et al., 2014; Riedel et al., 2015). Participants saw a written word on a screen and subsequently viewed a video without audio stream of a male speaker articulating a word. The articulated word was either the same as the previously presented written word or slightly altered version of the word (pseudoword). Participants indicated via button press whether the written word and the spoken word were the same or not.

In addition, in the context of a different research question, we assessed face recognition abilities using standard tests.

2.3. Eye tracking

During fMRI data acquisition, we recorded participants' eye movements using a 120 Hz monocular MR compatible eye tracker (EyeTrac 6, ASL, USA). The optical path was reflected over a mirror placed on top of the head coil in order to capture the image of the eye. Prior to the experiment, the eye tracking system was calibrated using a standard nine-point calibration procedure. The accuracy of eye tracking was checked before each run in the experiments. If necessary, the eye tracking system was recalibrated.

2.4. Image acquisition

Functional and structural data was acquired on a SIEMENS MAGNETOM Prisma (3 Tesla magnetic resonance imaging scanner (Siemens, Germany)). Functional images were collected with a 20-channel head coil using a gradient echo EPI (echo planar imaging) sequence (TR = 2790 ms, TE = 30 ms, flip angle = 90°, 42 slices, whole brain coverage; slice thickness = 2 mm; interslice gap = 1 mm; in-plane resolution = 3 × 3 mm).

A structural image was acquired using a 32-channel head coil and a T1-weighted 3D magnetization-prepared rapid gradient echo (MPRAGE) sequence (TR = 2300 ms; TE = 2.98 ms; TI = 900 ms; flip angle = 9°; FOV = 256 mm × 240 mm; voxel size = 1 mm^3 (isotropic resolution) 176 sagittal slices). This was done only for participants (n = 10) for whom no data was available from previous studies conducted at the Max Planck Institute for Human Cognitive and Brain Sciences in Leipzig. We accessed MPRAGE images available in the institute's data bank, which had been acquired also with a 32-channel coil and with the exact same acquisition parameters on 3 Tesla MRI scanners (SIEMENS MAGNETOM Trio, Verio and Prisma (Siemens, Germany).

2.5. Data analysis

2.5.1. Behavior

Behavioral data was analyzed with PASW Statistics 22.0 (IBM SPSS Statistics, USA). We computed group comparisons using analyses of variance (ANOVA) and Welch's independent samples *t*-test. Within-group comparisons were calculated with one-sample *t*-test and paired-samples *t*-tests. All *t*-tests were calculated two-tailed. Level of significance for all tests was defined at $\alpha = 0.05$. To estimate the effect sizes we used η^2 (Eta squared) and Cohen's *d*.

2.5.2. Eye tracking

Eye tracking data was analyzed offline (ASL Results Plus, Applied Science Laboratories, Bedford, USA). Data from 12 ASD participants and 12 control participants was included in the eye tracking data analysis. We had to exclude eye tracking data from the other participants due to difficulties with obtaining the corneal reflection (4 ASD and 1 control participant). Eye tracking data from their respective matched participants was also excluded.

A fixation was defined as having a minimum duration of 100 ms and a maximum visual angle change of 1°. For each participant, we measured the total number of fixations for the two conditions of the visual-speech recognition experiment (visual-speech task, face-identity task), and for the face conditions of the ROI localizer (static face, facial speech movement). To account for differences in the amount of successfully recorded data points due to inter-individual variance in trackability of pupil and/or corneal reflection, we normalized the number of fixations with a coefficient taking into account the relative duration of data loss: [(total data duration – duration of data loss)/total data duration].

To investigate where participants looked we created rectangular areas of interest (AOIs). For the visual-speech recognition experiment, we defined three AOIs: “Eye-AOI”, “Mouth-AOI” and “Off-AOI”. The first AOI covered the eyes (“Eye-AOI”): the left boundary of the rectangle was located 80 pixels to the left of the left pupil, the right boundary 80 pixels to the right of the right pupil, the upper boundary 60 pixels above the pupils, and the lower boundary 60 pixels below the pupils. The second AOI covered the mouth (“Mouth-AOI”): the left and right boundaries of the rectangle were located 110 pixels left and right of the center of the mouth, the upper and lower boundaries 60 pixels above and below the center. Fixations falling outside the AOIs “Eye” and “Mouth” were labeled as “Off-AOI”.

For the ROI localizer, we defined three AOIs: “Eye-AOI”, “Mouth-AOI” and “Off-AOI”. The first AOI covered the eyes (“Eye-AOI”): the left boundary of the rectangle was located 175 pixels to the left and the right boundary 175 pixels to the right from the middle point between the eyes, the upper boundary 45 pixels above and the lower boundary 45 pixels below the point. The second AOI covered the mouth (“Mouth-AOI”): the left and right boundaries of the rectangle were located 110 pixels left and right of the center of the mouth, the upper and lower boundaries 60 pixels above and below the center. Fixations falling outside the AOIs “Eye” and “Mouth” were labeled as “Off-AOI”.

We compared the total number of fixations between the groups and between the conditions using a repeated measures ANOVA. We defined 2 × 2 ANOVAs for both the visual-speech recognition experiment [*Task* (visual-speech, face-identity) × *Group* (control, ASD)] and the ROI localizer [*Movement* (static face, facial speech movement) × *Group* (control, ASD)]. Next, we looked at fixations onto the AOIs. For the visual-speech recognition experiment, we defined a 2 × 3 × 2 ANOVA: [*Task* (visual-speech, face-identity) × *AOI* (Eye, Mouth, Off) × *Group* (control, ASD)]. For the ROI localizer, we defined a 2 × 3 × 2 ANOVA: [*Movement* (static face, facial speech movement) × *AOI* (Eye, Mouth, Off) × *Group* (control, ASD)].

2.5.3. fMRI Analysis

2.5.3.1. Preprocessing and movement artifact correction. MRI data was analyzed using Statistical Parametric Mapping (SPM 12; Wellcome Trust Centre of Imaging Neuroscience, London, UK; http://www.fil.ion.ucl.ac.uk/spm) in a Matlab environment (version 10.11, The MathWorks, Inc., MA, USA). T2*- weighted images were spatially preprocessed using standard procedures: realignment and unwarp, normalization to Montreal Neurological Institute (MNI) standard stereotactic space using the T1 scan of each participant, smoothing with an isotropic Gaussian filter of 8 mm at FWHM, and high-pass filtering at 128 s. Geometric distortions due to susceptibility gradients were corrected by an interpolation procedure based on the B0 field-map (Jezzard and Ballaban, 1995).

To control for potential confounding effects of movement artefacts on the BOLD signal change we examined the head movement along six possible axes during both experiments. We compared 6 movement parameters resulting from rigid body transformation during spatial

realignment using independent-samples *t*-test. For both experiments, we found significant group differences in head movement along three axes (translation along x-axis, rotation around yaw and rotation around roll) indicating that the ASD group moved significantly more than the control group (Supplementary Table S1). Such finding is in accordance with previous literature (for a review see Travers et al. 2012). In order to control for the movement differences between the groups, we examined each participant's functional time series for global-signal artefacts using the Artifact Detection Tool (ART) software package (http://web.mit.edu/swg/art/art.pdf). Volumes were flagged as "outlier" volumes if the average global-signal intensity of the image (i.e., average signal intensity across all voxels) was more than 3.0 standard deviations from the overall mean for all images (ART z-threshold = 3.0), and the absolute global translation movement was more than 3 mm. Outlier volumes and 6 movement parameters were modeled as covariates of no interest in the first-level GLM. There were no significant group differences in the number of outlier volumes in any of the two experiments (Supplementary Table S1).

2.5.3.2. BOLD response analysis. At the first level, statistical parametric maps were generated by modeling the evoked hemodynamic response for the different conditions as boxcars convolved with a synthetic hemodynamic response function in the context of GLM (Friston et al., 2007). For the ROI localizer, we modeled the conditions "static face", "facial speech movement", "static objects" and "object movement". Head movement parameters and outlier volumes were included as covariates of no interest. For the visual-speech recognition experiment, we modeled the conditions "visual-speech task", "face-identity task" and "instruction". Head movement parameters and outlier volumes were modeled as covariates of no interest. To account for potential effects of eye movements on brain responses, eye tracking data was also included into the first-level analysis. We entered the normalized number of eye fixations onto the predefined AOIs ("Eye-AOI", "Mouth-AOI", "Off-AOI") in the visual-speech task and the face-identity task as three covariates of no interest (except for the participants for whom this data was not available, see section "Eye tracking").

At the second-level, population-level inferences about BOLD response changes were based on a random effects model that estimated the second-level statistic at each voxel. For the ROI localizer and the visual-speech recognition experiment, we performed one-sample *t*-tests across the single-participant contrast images as within-group analyses. For between-group analyses, we used two-sample *t*-tests comparing the means of the single-subject contrast images from both groups. For the visual-speech recognition experiment, we included the difference between correct responses in the visual-speech task and in the face-identity task as a covariate of no interest to control for different difficulty levels of the two tasks.

2.5.3.3. Correlation analysis. To further assess behavioral relevance of BOLD response to visual-speech recognition, we performed correlation analyses using SPM12. To do this, we entered visual-speech recognition performance score as a covariate of interest into the second-level analysis. This was done separately for the fMRI visual-speech task score and the visual word-matching test score. We do not report the *r* values as an estimate of the effect size of a correlation, because SPM does not provide *r* values.

2.5.3.4. ROI definition. The ROIs included: the visual-movement regions (bilateral V5/MT and left TVSA, Fig. 2A), and the speech-motor regions (left IFG, bilateral PCG and bilateral SMA; Fig. 2B). The choice of the regions was based on findings in previous literature (Blank and von Kriegstein, 2013; Bernstein and Yovel, 2015).

Visual-movement ROIs in the bilateral V5/MT and in the left TVSA were defined by means of the ROI localizer. The right and the left V5/MT were localized based on the contrast "(facial speech movement + object movement) > (static face + static object)", because the V5/MT region is known to be involved in perception of both human and non-human movement (Zeki et al., 1991; Grèzes et al., 2001). The left TVSA was localized using the contrast "facial speech movement > static face", because it is known to be relevant for human-only movement processing including visual speech (Puce et al., 1998; Grossman et al., 2005; Bernstein et al., 2011). We adopted the term "temporal visual speech area (TVSA)" from Bernstein et al. (2011) to refer to the portions of the left posterior STS/STG that were sensitive to the facial speech movement compared to the static face condition. Initially, the TVSA was defined using a different contrast between speech and non-speech facial movements (visual-speech > visual non-speech) ∩ (point-light speech > point-light non-speech; Bernstein et al., 2011). The TVSA definition in our study might contain also other regions compared to TVSA by Bernstein et al. (2011), because our control condition "static face" included only the face, but no movement. We defined the visual-movement ROIs both on the group level and in each individual participant. The speech-motor ROIs were defined only at the group level.

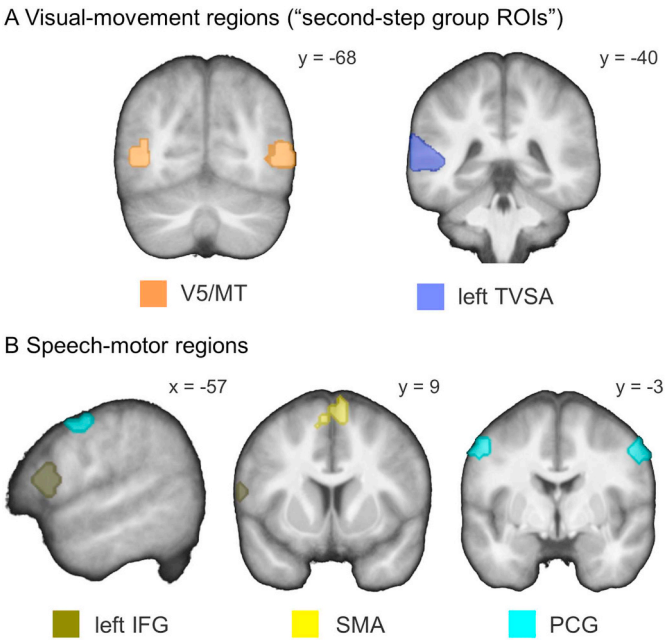

Fig. 2. Regions of interest for the second-level ROI analysis of the visual-speech recognition experiment. **(A)** Visual-movement regions were localized functionally on the group level using the ROI localizer ("second-step group ROIs"). BOLD response to the contrast "(facial speech movement + object movement) > (static face + static object)" was localized in the bilateral V5/MT. Contrast "facial speech movement > static face" elicited BOLD response in the left pSTS/STG labeled as the temporal visual speech area (TVSA). **(B)** Speech-motor regions were defined based on probabilistic anatomical maps of the Harvard-Oxford cortical structural atlas (Desikan et al., 2006). All ROIs are overlaid onto a sample specific average image of normalized T1-weighted structural images of all participants in the study (n = 34). V5/MT = visual area 5/middle temporal area; TVSA = temporal visual speech area; pSTS/STG = posterior superior temporal sulcus/ gyrus; IFG = inferior frontal gyrus; SMA = supplementary motor area; PCG = precentral gyrus. x, y = MNI coordinates.

2.5.3.4.1. Group-level ROIs. The group-level ROIs were defined for the purpose of a second-level ROI analysis of BOLD response and functional connectivity in the visual-speech recognition experiment.

2.5.3.4.1.1. Visual-movement regions

The ROIs in the bilateral V5/MT and the left TVSA were defined by a combined functional and anatomical approach. The peak coordinates for the ROIs were first defined functionally based on BOLD response to

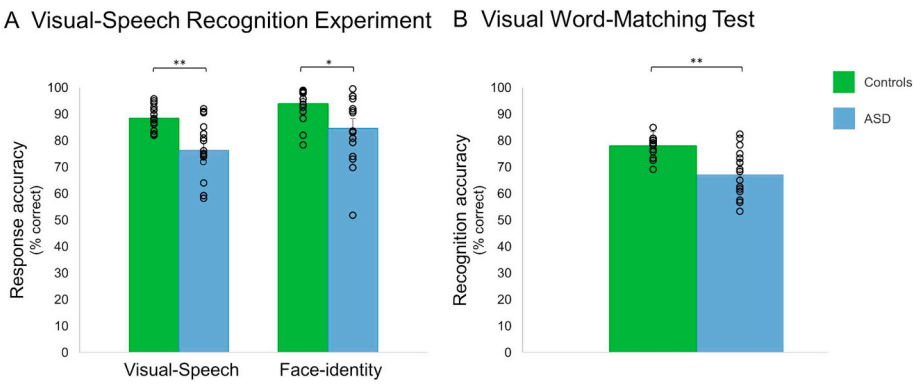

Fig. 3. Behavioral performance of the ASD group and the control group in tests on visual-speech and face-identity recognition measured in % correct recognition. **(A)**Visual speech recognition experiment (during fMRI): The ASD group was significantly worse in recognizing visual speech and face identity compared to the control group **(B)** Visual-word matching test: The ASD group was significantly worse in matching silently articulated and written three-syllabic words in comparison to the control group. We display individual (i.e. circles) and mean-group (i.e. bars) results. For the exact individual participant values, see Supplementary Table S4. Error bars represent +/− 1 SE; $^{*}p < .05$; $^{**}p < .001$.

Table 2
Summary of normalized number of eye fixations during visual-speech recognition experiment and ROI localizer.

	Control (n = 12)		ASD (n = 12)			
	M	*SD*	*M*	*SD*	*p*	*d*
Visual-speech recognition experiment						
Visual-speech task						
Total	520.58	115.26	533.67	62.13	0.733	0.141
Eyes	177.33	55.81	184.92	44.68	0.717	0.150
Mouth	195.92	64.03	186.75	40.46	0.680	0.171
Off	147.33	37.21	161.33	18.79	0.262	0.475
Face-identity task						
Total	520.83	85.53	530.67	90.61	0.787	0.111
Eyes	223.33	47.66	219.42	36.41	0.824	0.092
Mouth	157.75	42.00	150.50	49.30	0.702	0.158
Off	139.75	32.25	160.42	40.08	0.178	0.568
ROI localizer						
Facial-speech movement						
Total	203.58	51.57	214.58	58.69	0.631	0.199
Eyes	78.42	38.83	91.83	38.94	0.407	0.345
Mouth	34.17	22.76	37.09	34.96	0.811	0.099
Off	91.00	30.85	85.67	26.11	0.652	0.186
Static face						
Total	190.42	53.63	213.17	59.72	0.337	0.401
Eyes	103.92	38.15	89.42	41.29	0.381	0.364
Mouth	11.25	24.04	28.00	38.17	0.212	0.525
Off	75.25	34.88	95.75	34.52	0.162	0.591

M = mean; SD = standard deviation; d = Cohen's d

the respective contrasts of interest in the ROI localizer. The ROI localizer was thresholded at $p < .05$ uncorrected. The clusters were then masked with a probabilistic anatomical map of the respective brain region implemented in FSL (V5/MT: Jülich histological (cyto- and myelo-architectonic) atlas (Eickhoff et al., 2007); the left pSTS/STG for the TVSA: Harvard-Oxford cortical structural atlas (Desikan et al., 2006); FSL (Smith et al., 2004), http://www.fmrib.ox.ac.uk/fsl/fslview). Because the anatomical map of the left pSTS/STG partially overlaps with the anatomical map of the left anterior STS/STG (aSTS/STG), we subtracted the left aSTS/STG map from the left pSTS/STG map, to ensure that the anatomical map of the left pSTS/STG does not contain any aSTS/STG regions. The overlap between the functional and the anatomical maps was defined as the bilateral V5/MT and the left TVSA ROIs. We chose the combined functional and anatomical approach for ROI definition to ensure that the V5/MT and the TVSA ROIs (i) do not overlap with each other, and (ii) are restrained to regions that have been anatomically predefined as the V5/MT and the pSTS/STG for the TVSA.

In a first step, we created the ROIs based on BOLD response of all the participants from both groups ("first-step group ROIs"; Supplementary Fig. S1A). However, when we looked on each group separately, this approach did not allow us to define the left V5/MT in the ASD group (see Supplementary Fig. S1A). In a next step, we therefore defined the visual-movement ROIs based on BOLD response of only those participants of both groups who showed a significant BOLD response ($p < .09$ uncorrected) in the respective brain region on the single-participant level (Supplementary Table S2). We will call these ROIs "second-step group ROIs". The number of the included participants per group varied between the regions (right V5/MT: $n_{CON} = 12$ and $n_{ASD} = 12$; left V5/MT: $n_{CON} = 11$ and $n_{ASD} = 11$; left TVSA: $n_{CON} = 15$ and $n_{ASD} = 11$). With this approach, we were able to localize all the three visual-movement regions in each group separately (Supplementary Fig. S1B). We used the "second-step group ROIs" to define the visual-movement ROIs for the second-level ROI analysis of the visual-speech recognition experiment (Fig. 2A). We also performed control analyses with the "first-step group ROIs" to check whether the reported effects are robust to different ROI definition approaches.2.5.3.4.1.2. Speech-motor regions

Speech-motor ROIs were defined using mean coordinates of the respective brain regions that were previously reported by Blank and von Kriegstein (2013). This study used a very similar design to our study, and contrasted tasks on visual-speech and face-identity recognition performed on identical stimulus material (Blank and von Kriegstein, 2013). We created spheres of 8 mm around MNI coordinates of the speech regions. To ensure that the spheres were located within the anatomically defined left IFG, bilateral PCG and bilateral SMA, we masked the spheres with probabilistic anatomical maps of the respective brain regions from the Harvard-Oxford cortical structural atlas (Desikan et al., 2006) implemented in the FSL software (Smith et al., 2004, http://www.fmrib.ox.ac.uk/fsl/fslview). The overlaps between the spheres and the anatomical maps and were defined as the respective ROIs (Fig. 2B).

2.5.3.4.2. Individual ROIs: Visual movement regions. We defined individual ROIs in the visual-movement regions for the purpose of defining seed regions in the functional connectivity analysis and for a control analysis of BOLD response on the single-participant level. We used the following procedure: For each participant, we identified the three visual-movement regions (i.e. right V5/MT, left V5/MT, left TVSA) defined as 4 mm-radius spheres centered on their individual peak responses obtained from the respective contrasts of interest in the ROI localizer (Supplementary Table S2). If there was no peak in the individual participant even at a lenient threshold ($p < .09$ uncorrected to reduce type II error, i.e. missing an individual participant's peak), we used the group coordinate from the contrast of interest (see Section 2.5.3.4.1 Group-level ROIs). To ensure that the individual spheres were located within the anatomically defined V5/MT and pSTS/STG for the TVSA, the 4 mm-radius spheres were overlaid with a probabilistic anatomical mask of the respective brain region implemented in FSL (V5/MT: Jülich histological (cyto- and myelo-architectonic) atlas (Eickhoff et al., 2007); the left pSTS/STG for the TVSA: Harvard-Oxford cortical structural atlas (Desikan et al., 2006)). Again, we

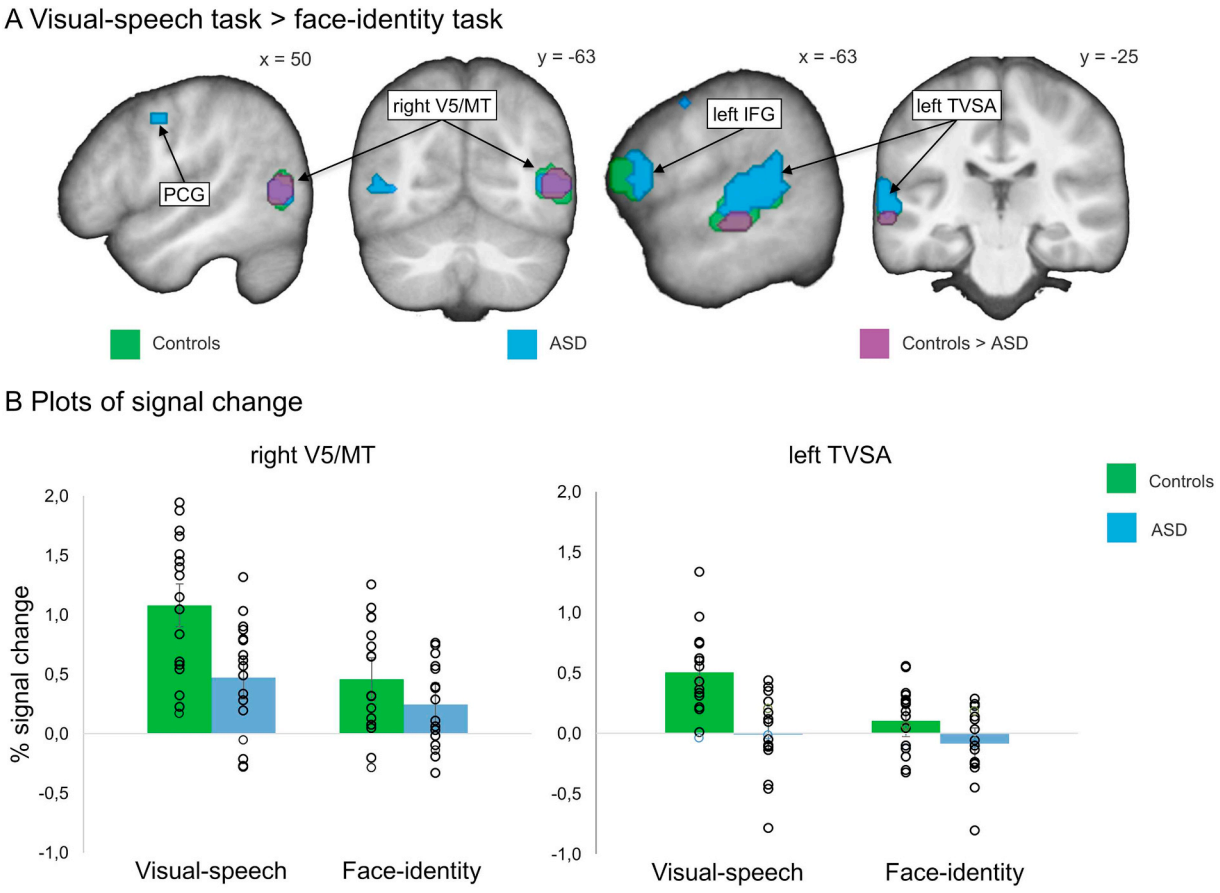

Fig. 4. Results of the visual-speech recognition experiment. **(A)** The contrast visual-speech task > face-identity task is shown for the control group (green), and the ASD group (blue). The interaction between *Task* (visual-speech, face-identity) and *Group* (control, ASD) is shown in purple. For display purposes only, within-group effects are presented at the threshold of $p = .001$ uncorrected, and between-group effects are presented at the threshold of $p = .05$ (same masks as for ROI analyses). All results were overlaid onto a sample specific average image of normalized T1-weighted structural images. **(B)** Percent signal change for each condition separately extracted at the maximum statistic for the *Task* × *Group* interaction. We display individual (i.e. circles) and mean-group (i.e. bars) results. For the exact individual participant values, see Supplementary Table S4. V5/MT = visual area 5/middle temporal area; TVSA = temporal visual speech area; IFG = inferior frontal gyrus; PCG = precentral gyrus; SMA = supplementary motor area. x, y, MNI-coordinates. Error bars represent +/− 1 SE.

subtracted the left aSTS/STG map from the left pSTS/STG map, to ensure that the anatomical map of the left pSTS/STG does not contain any anterior STS/STG regions. The overlap between the spheres and the anatomical maps was defined as the individual visual-movement ROIs. Volumes of the individual visual-movement ROIs were not significantly different between the control and the ASD group ($p \geq .100$, Supplementary Table S3).

2.5.3.5. Functional Connectivity (Psychophysiological Interactions, PPI). We investigated functional connectivity during the visual-speech task compared to the face-identity task between: (i) visual-movement regions, and (ii) visual-movement regions and speech regions. Functional connectivity was assessed by psycho-physiological interaction (PPI) analysis using routines implemented in SPM12 (Friston et al., 1997). The seed regions were defined in the right V5/MT, in the left V5/MT and in the left TVSA. These seed regions were identified in each individual participant by finding the peak of the contrast visual-speech task > face-identity task that was located within the respective individual visual-movement ROI that we predefined using the ROI localizer (see section 2.5.3.4.2 Individual ROIs: visual-movement regions). The first Eigenvariate was extracted from the respective seed regions in each individual participant. The psychological variable was the contrast "visual-speech task > face-identity task". At the first level, the PPI regressor, the psychological variable, and the first Eigenvariate were entered as covariates into a design matrix. At the second level, we performed one-sample *t*-tests across the single-subject contrast images as within-group analyses. For between-group analyses, we used two-sample *t*-tests comparing the means of the single-subject contrast images from both groups. Population-level inferences about BOLD response changes were based on a random effects model that estimated the second-level statistic at each voxel.

2.5.3.6. Significance thresholds for fMRI second-level analyses. For the BOLD response analysis and functional connectivity analysis, effects were considered significant at $p < .05$ corrected for family wise error (FWE) for the ROI (i.e. small volume correction). The ROIs included three visual-movement regions (bilateral V5/MT and left TVSA, Fig. 2A), and five speech-motor regions (left IFG, bilateral PCG and bilateral SMA; Fig. 2B). The visual movement ROIs were the "second-step group ROIs" (see Section 2.5.3.4.1 Group-level ROIs). We applied the Holm-Bonferroni method to correct for multiple comparisons for the eight ROIs (right V5/MT, left V5/MT, left TVSA, left IFG, right PCG, left PCG, right SMA, left SMA) (Holm, 1979). We chose this method because it is considered a conservative method for multiple comparisons, and it is less susceptible to Type II error (i.e. missing true effects) in comparison to the standard Bonferroni correction (Nichols and Hayasaka, 2003). Other effects outside the ROIs were considered significant at $p < .05$ FWE corrected for the whole brain.

2.5.3.7. Control analyses. We conducted three control analyses. First, we conducted a BOLD response analysis on the single-participant level

Table 3
Coordinates of brain areas showing significant BOLD response in the visual-speech recognition experiment.

Visual-speech task > face-identity task									
Region		x	y	z	*Z*	x	y	z	*Z*
		Control				ASD			
V5/MT	r	57	−64	2	5.43	48	−61	5	3.92
						54	−67	2	3.91
	l	−48	−64	5	4.54	−48	−64	8	4.84
		−39	−70	8	4.20				
TVSA	l	−57	−34	14	5.39	−57	−34	11	4.75
		−60	−28	−1	5.13	−66	−25	11	4.75
		−60	−31	8	5.01	−60	−40	8	4.70
		−63	−16	−7	4.78	−63	−31	8	4.66
		−66	−37	23	4.38				
		−63	−34	20	4.35				
		−66	−43	5	4.07				
IFG	l	−54	17	11	4.84	−57	8	14	4.17
						−54	14	17	4.09
PCG	r	51	−1	44	4.62	57	−1	41	5.73
	l	−48	−1	44	4.77	−54	−7	44	5.04
		−51	2	41	4.62				
SMA	r	0	5	68	4.06	3	2	65	5.47
		6	2	68	4.02				
	l	−3	8	62	3.87	−3	8	59	4.77
		Control > ASD				ASD > Control			
V5/MT	r	54	−64	5	3.46	–			
TVSA	l	−63	−25	−7	4.04	–			

Coordinates represent local response maxima in MNI space (in mm). Clusters are reported that reached significance at $p = .006$ FWE corrected (peak-level) for the respective ROI and Holm-Bonferroni corrected for eight ROIs, and which cluster size contained more than 5 voxels. Regions were labeled using a standard anatomical atlas (Harvard-Oxford cortical and subcortical structural atlases (Desikan et al., 2006) and Jülich histological (cyto- and myelo-architectonic) atlas (Eickhoff et al., 2007)), implemented in FSL (Smith et al., 2004; http://www.fmrib.ox.ac.uk/fsl/fslview). V5/MT = visual area 5/middle temporal area; TVSA = temporal visual speech area; IFG = inferior frontal gyrus; PCG = precentral gyrus; SMA = supplementary motor area; *Z* indicates the statistical value.

to check whether the results obtained with the group ROIs were reproducible for ROIs defined on the single-participant level. We did this because individual variability in spatial location of the motion-sensitive V5/MT region and the TVSA is known (e.g., Allison et al., 2000; Watson et al., 1993; Malikovic et al., 2006). Second, we repeated the second-level ROI analysis using the "first-step group ROIs" in the visual-movement regions to check whether the potential effects in these regions can be replicated with a different ROI definition approach that comprises the whole participant sample. Third, we investigated whether potential functional alterations in the visual-movement regions are specific to the motion-sensitive visual areas or also present in other early visual areas. For more details see Supplementary Methods.

3. Results

3.1. Recognition of visual speech is impaired in ASD

For the visual-speech recognition fMRI experiment, a repeated measures ANOVA with the within-subject factor *Task* (visual-speech, face-identity) and the between-subject factor *Group* (control, ASD) revealed that performance for both tasks, i.e. visual-speech task and face-identity task, was impaired in the ASD group compared to the control group ($F(1,32) = 14.469$, $p = .001$, $\eta^2 = 0.311$) (Fig. 3A). Both groups performed significantly better in the face-identity task compared to the visual-speech task ($F(1,32) = 25.523$, $p = .000$, $\eta^2 = 0.444$). The group differences were confirmed in a post-hoc analysis using Welch's *t*-tests (visual-speech task: $t(21) = 4.221$, $p = .000$, $\eta^2 = 1.444$; face-identity task: $t(23) = 2.769$, $p = .011$, $\eta^2 = 0.950$) (Fig. 3A). There was no *Task* × *Group* interaction ($p = .315$).

For the visual word-matching test, a Welch's *t*-test revealed significant group differences where the ASD group performed worse than the control group ($t(21) = 4.553$, $p = .000$, Cohen's $d = 1.562$) (Fig. 3B).

3.2. ASD and control participants showed similar gaze behavior

3.2.1. Visual-speech recognition experiment

For the normalized total number of fixations, a repeated measures ANOVA with the within-subject factor *Task* (visual-speech, face-identity) and the between-subject factor *Group* (control, ASD) showed no significant group effects. This indicates that in both tasks, the ASD group and the control group fixated the videos with a similar frequency (both $p \geq .337$; Table 2).

To investigate fixation patterns to the different AOIs, we conducted a repeated measures ANOVA with within-subject factors *Task* (visual-speech, face-identity) and *AOI* (Eye, Mouth, Off), and the between-subject factor *Group* (control, ASD). There was a significant interaction between *Task* × *AOI* ($F(2,21) = 10.062$, $p = .001$, $\eta^2 = 0.489$), indicating that participants looked at different regions of the face during the two tasks. This adaptation of gaze behavior to task demands was similar for the two groups as the three-way interaction *Task* × *AOI* × *Group* was not significant ($F(2,21) = 0.171$, $p = .844$, $\eta^2 = 0.016$). To further investigate the cause of the interaction *Task* × *AOI* for the visual-speech recognition experiment, we repeated the analysis for each task separately with the within-subject factor *AOI* (Eye, Mouth, Off), and found a main effect of AOI for both tasks (visual-speech task: $F(2,21) = 7.004$, $p = .005$, $\eta^2 = 0.400$; face-identity task: $F(2,21) = 34.002$, $p = .000$, $\eta^2 = 0.764$). Post-hoc *t*-tests conducted for each task separately showed that in the visual-speech task, participants fixated significantly more to the Eye-AOI ($t(23) = 2.300$, $p = .031$, Cohen's $d = 0.655$), and to the Mouth-AOI ($t(23) = 3.567$, $p = .002$, Cohen's $d = 0.866$) compared to the Off-AOI. There was no significant difference between the Eye-AOI and the Mouth-AOI ($t(23) = -0.766$, $p = .451$, Cohen's $d = 0.200$). In the face-identity task, participants fixated significantly more to the Eye-AOI compared to the Mouth-AOI ($t(23) = 5.744$, $p = .000$, Cohen's $d = 1.554$), and compared to the Off-AOI ($t(23) = 8.108$, $p = .000$, Cohen's $d = 1.810$). There was no difference between the Mouth-AOI and the Off-AOI ($t(23) = 0.377$, $p = .710$, Cohen's $d = 0.098$) (Table 2).

3.2.2. ROI localizer

We performed a repeated-measures ANOVA for the normalized number of fixations with the within-subject factor *Movement*(static face, facial speech movement) and the between-subject factor *Group* (control, ASD). No significant main effects of *Movement* and *Group* were found indicating that ASD and control group fixated the images with a similar frequency for both conditions (Table 2).

Furthermore, we analyzed fixation patterns for the AOIs by conducting a repeated measures ANOVA with the within-subject factors *Movement* (static face, facial speech movement) and *AOI* (Eye, Mouth, Off), and the between-subject factor *Group* (control, ASD). The two-way interaction between *Movement* and *AOI* ($F(2,21) = 2.796$, $p = .067$, $\eta^2 = 0.295$), and the three-way interaction *Movement*, *AOI* and *Group* did not reach significance ($F(2,21) = 2.303$, $p = .108$, $\eta^2 = 0.257$). The results show that fixation behavior in both face conditions was similar for the ASD and the control group (Table 2).

Altogether, the findings showed that the fixation behavior in both visual-speech recognition experiment and ROI localizer was remarkably similar in the ASD group and the control group.

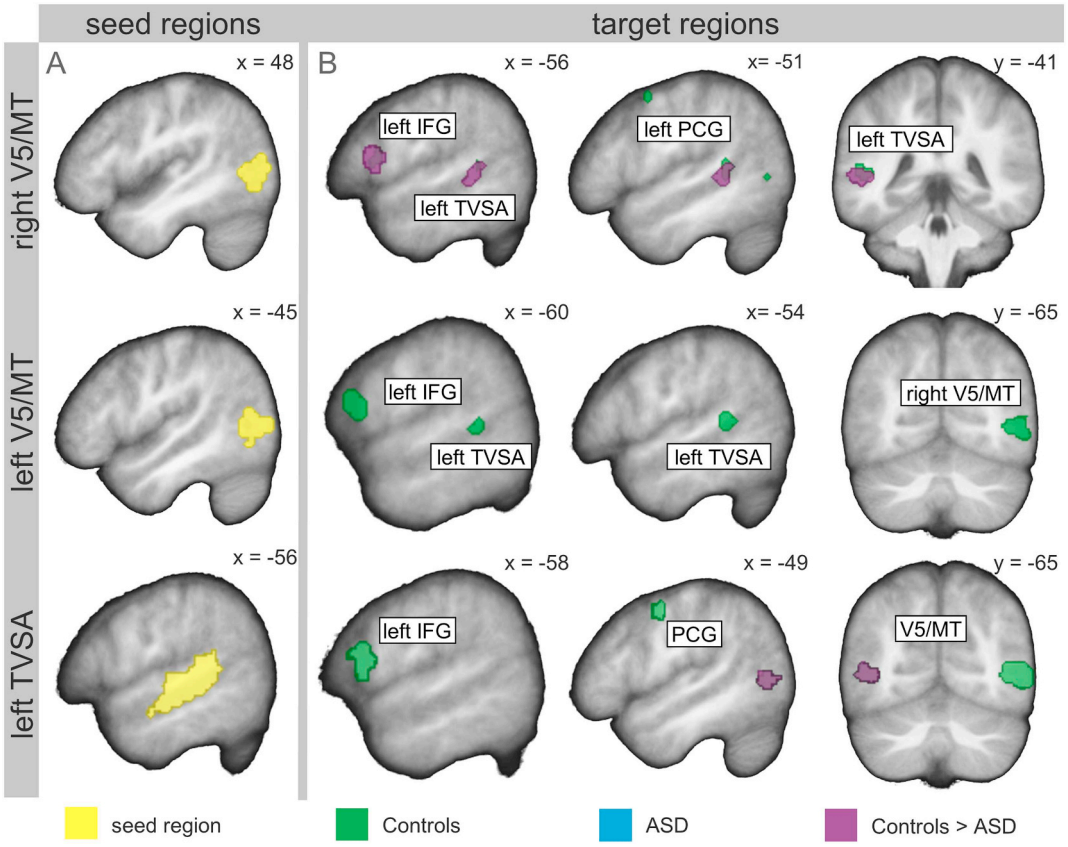

Fig. 5. Functional connectivity of seed regions in the bilateral V5/MT and in the left TVSA during visual-speech recognition. **(A)** Seed regions were located within a radius of 4 mm from the individual maximum of the bilateral V5/MT and the TVSA response that was defined by the ROI localizer in each single participant. Regions marked in yellow illustrate the bilateral V5/MT and the left pSTS/STG for the TVSA region within which individual seed regions were located. **(B)** The bilateral V5/MT and the left TVSA were functionally connected to each other and to the speech-motor regions (left IFG; PCG; SMA) in the control group (green) and in the ASD group (blue). The control group showed higher functional connectivity than the ASD group between the left TVSA and the bilateral V5/MT, and between the right V5/MT and the left IFG (purple). For display purposes within-group effects are presented at the threshold of $p = .001$ uncorrected, and between-group effects are presented at the threshold of $p = .05$ (same masks as for ROI analyses). All results are overlaid onto a sample specific average image of normalized T1-weighted structural images. V5/MT = visual area 5/middle temporal area; TVSA = temporal visual speech area; IFG = inferior frontal gyrus; PCG = precentral gyrus; SMA = supplementary motor area. x, y, MNI-coordinates.

3.3. Brain responses to visual-speech vs. face-identity task are reduced in visual-movement regions, but not in other speech regions in ASD

Both groups showed significantly higher BOLD response in the bilateral V5/MT and in the left TVSA when performing the visual-speech task compared to the face-identity task ($p \leq .002$ FWE corrected within the ROI, Holm-Bonferroni corrected; Fig. 4A; Table 3). Similarly in the speech-motor regions, both groups showed increased BOLD response in the left IFG, and in the bilateral PCG and SMA (all $p \leq .001$ FWE corrected, Holm-Bonferroni corrected) (Fig. 4A; Table 3).

In the right V5/MT and in the left TVSA, we identified a significant *Task* × *Group* interaction: [(visual-speech task/control > face-identity task/control) > (visual-speech task/ASD > face-identity task/ASD)] (Fig. 4A; Table 3). The interaction in both regions remained significant after Holm-Bonferroni correction for the eight ROIs (right V5/MT: x = 54, y = −64, z = 5; $p = .007$ FWE corrected, within the ROI; left TVSA: x = −63, y = −25, z = −7; $p = .003$ FWE corrected, within the ROI). Extracting the percent signal change from the second-level analysis for each condition separately suggested that the interaction was caused by BOLD response differences between the control group and the ASD group in the visual-speech task, rather than the face-identity task (Fig. 4B). For responses in the left V5/MT and in the three speech regions, there was no significant *Task* × *Group* interaction (all $p > .018$ uncorrected, Supplementary Fig. S2). The group differences in BOLD response are unlikely to be primarily caused by gaze behavior or head movement, as we included all the parameters as covariates of no interest into the analysis. In addition, the *Task* × *Group* interaction (Fig. 4A, purple) is unlikely due to behavioral differences between the groups, as there was no *Task* × *Group* interaction at the behavioral level (see Fig. 3A).

3.4. Reduced functional connectivity between visual-movement regions in ASD

The bilateral V5/MT and the left TVSA (Fig. 5A) were functionally connected to each other and to the speech-motor regions (left IFG, PCG, SMA) in the control group (Fig. 5B, green; Table 4), and in the ASD group (Fig. 5B, blue; Table 4). All effects were significant at $p < .05$ FWE corrected (within the ROI) and remained significant after applying the Holm-Bonferroni correction.

Group comparison revealed higher functional connectivity between the visual-movement regions in the control group compared to the ASD group, i.e. between the left TVSA and the left V5/MT ($p = .007$ FWE corrected, within the ROI, Holm-Bonferroni corrected), and between the right V5/MT and the left TVSA ($p = .008$ FWE corrected, within the ROI, Holm-Bonferroni corrected; Fig. 5B, purple; Table 4). Functional connectivity between the visual-movement regions and the speech-motor regions was significantly higher between the right V5/MT and the left IFG in the control group compared to the ASD group ($p = .007$ FWE corrected, within the ROI, Holm-Bonferroni corrected; Fig. 5B,

Table 4
Coordinates for brain areas showing functional connectivity to the bilateral V5/MT and to the left TVSA during visual-speech recognition.

Region		x	y	z	*Z*	x	y	z	*Z*
Seed region: right V5/MT									
		Control				ASD			
V5/MT	l	−45	−73	5	4.86	−39	−73	2	3.40
		−33	−76	2	3.43				
TVSA	l	−51	−46	8	3.88	–			
		−48	−43	5	3.67				
IFG	l	−54	11	11	3.73	–			
PCG	l	−51	2	50	3.65	–			
SMA	l	−3	8	59	4.17	–			
	r	0	5	68	3.04				
		6	8	71	3.01				
		Control > ASD				ASD > Control			
TVSA	l	−48	−43	2	3.71	–			
IFG	l	−54	11	11	3.35	–			
Seed region: left V5/MT									
		Control				ASD			
V5/MT	r	51	−64		−7	3.64	–		
TVSA	l	−57	−40	5	2.98	–			
IFG	l	−57	11	17	3.09	–			
PCG			–			–			
SMA	l	−9	8	56	3.04	–			
		Control > ASD				ASD > Control			
		–				–			
Seed region: left TVSA									
		Control				ASD			
V5/MT	r	48	−61	5	5.45	–			
		45	−70	−1	4.77				
		48	−67	−4	4.57				
	l	−45	−73	2	4.82	–			
IFG	l	−60	11	20	3.87	–			
PCG	l	−48	−1	47	4.61	–			
SMA	l	−9	8	56	3.25	–			
		Control > ASD				ASD > Control			
V5/MT	l	−48	−67	−1	3.45	–			

Coordinates represent local connectivity maxima in MNI space (in mm) for the whole brain. Clusters are reported that reached significance at $p = .007$ FWE corrected (peak-level) for the respective ROI and Holm-Bonferroni corrected for seven ROIs, and which cluster size contained more than 5 voxels. Anatomically, regions were labeled using a standard anatomical atlas (Harvard-Oxford cortical and subcortical structural atlases (Desikan et al., 2006) and Jülich histological (cyto- and myeloarchitectonic) atlas (Eickhoff et al., 2007)) implemented in FSL (Smith et al., 2004; http://www.fmrib.ox.ac.uk/fsl/fslview). V5/MT = visual area 5/ middle temporal area; TVSA = temporal visual speech area; IFG = inferior frontal gyrus; PCG = precentral gyrus; SMA = supplementary motor area; *Z* indicates the statistical value.

purple; Table 4). There was also a significant group difference for the functional connectivity between the left TVSA and the left IFG, but it did not remain significant after Holm-Bonferroni correction ($p = .036$ FWE corrected, within the ROI). For the remaining speech-motor regions, there were no significant group differences ($p > .027$ uncorrected).

3.5. Correlations with behavioral performance

In a next step, we tested whether local BOLD responses in the visual-movement regions that showed different responses between the groups during the visual-speech task vs. face-identity task (i.e. right V5/MT and left TVSA) were related to behavioral visual-speech recognition abilities. To do this, we computed correlations between local BOLD responses in the right V5/MT and in the left TVSA to the visual-speech task vs. face-identity task, and the behavioral performance assessed in the visual-speech task of the fMRI visual-speech recognition experiment and in the visual word-matching test. Hence, we computed four correlation calculations, for which we corrected our analysis ($p < .0125$ FWE corrected).

In the ASD group, BOLD response to the visual-speech task vs. face-identity task in the right V5/MT (x = 45, y = −64, z = 11) correlated positively with the visual-speech task performance ($p = .010$ FWE corrected, within the ROI; Fig. 6), and the visual word-matching test performance (x = 48, y = −67, z = 14, $p = .046$ FWE corrected, within the ROI). Only the first correlation remained significant after Holm-Bonferroni correction for the four tests ($p < .0125$ FWE corrected). In the control group, there was no correlation of the right V5/MT response with the visual-speech recognition performance. There were no correlations between behavioral performance and left TVSA

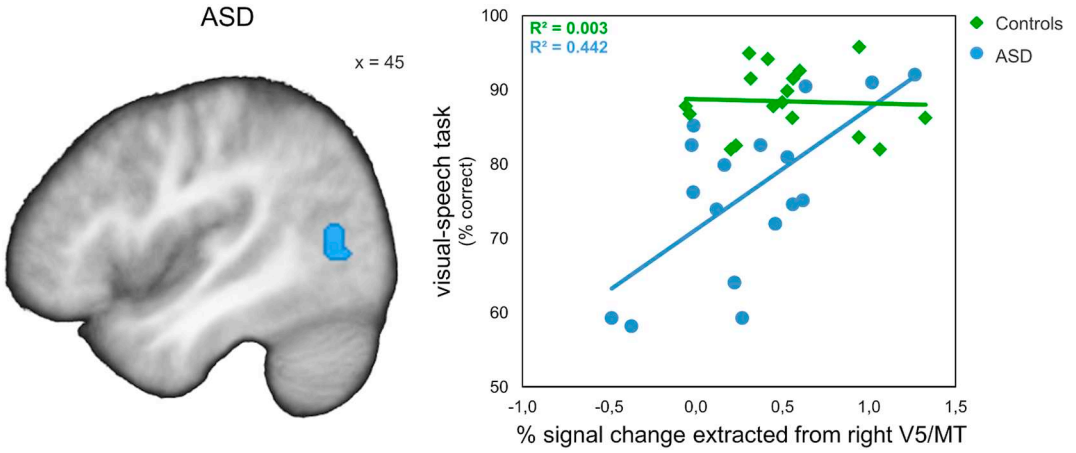

Fig. 6. Brain-behavior correlation. Behavioral performance in the visual-speech task correlated significantly positively with BOLD response in the right V5/MT in the ASD group (p = .010 FWE corrected, within the ROI, Holm-Bonferroni corrected), but not in the control group (p = .050 uncorrected, within the ROI). Correlation plot illustrates the relationship between the behavioral visual-speech recognition performance and the percent signal change extracted from the right V5/MT (x = 45, y = –64, z = 11) from the contrast visual-speech task > face-identity task. The lines for each group represent the best-fitting linear regression. V5/MT = visual area 5/ middle temporal area, x = MNI-coordinate.

responses in either of the groups (p > .085 FWE corrected for the respective ROI).

3.6. Control analyses

We conducted three control analyses. The results of the first two control analyses including the single-participant analysis and the second-level ROI analysis using "first-step group ROIs" in the visual-movement regions confirmed that BOLD responses to the visual-speech task (in comparison to the face-identity task) in the right V5/MT and in the left TVSA were higher in the control group compared to the ASD group. Again, we did not find any significant group differences in the left V5/MT.

In the bilateral early visual cortex, the third control analysis revealed that BOLD response to the visual-speech task in contrast to the face-identity task was not different between the groups, indicating that the reduced responses in the visual-movement regions in the ASD group are specific to visual-movement sensitive mechanisms and not general to the visual system of the ASD group. For more details, see Supplementary Results.

4. Discussion

The present study provided four key findings. First, adults with high-functioning ASD had reduced responses to recognition of visual speech, in contrast to face identity, in visual-movement regions (i.e. in the right V5/MT and the left TVSA). Second, the right V5/MT responses to visual speech were positively correlated with the performance in the visual-speech recognition task in the ASD group, but not in the control group. Third, the visual-movement regions were less functionally connected to each other and to the left IFG in the ASD group compared to the controls. Fourth, responses in the speech-motor regions (left IFG, bilateral PCG and SMA), and functional connectivity to the speech-motor regions (PCG, SMA) were at the level of the neurotypical participants. The results imply that a dysfunction at a perceptual level of visual-motion processing might underlie the impairment in visual-speech recognition in ASD, rather than difficulties at subsequent stages of speech analysis. This supports the view that at least part of the communication difficulties in ASD might stem from dysfunctional perceptual mechanisms (Baum et al., 2015; Herrington et al., 2007), and poses a challenge to accounts that attribute communication difficulties in ASD entirely to non-perceptual difficulties (e.g. Baron-Cohen, 1997; Chevallier et al., 2012).

Several previous neuroimaging studies in ASD have reported decreased V5/MT and pSTS/STG responses to human motion (Herrington et al., 2007; Freitag et al., 2008; Sato et al., 2012; Alaerts et al., 2013; Alaerts et al., 2017), as well as to non-human motion (Brieber et al., 2010; Robertson et al., 2014). Our study makes three important advances in relation to these studies. First, it is the first study to include a concurrent recording of eye tracking data during fMRI acquisition for visual-movement perception. The finding that the eye movements in the ASD group were relatively similar to controls implies that the V5/MT and the TVSA response differences to visual-speech recognition are unlikely due to differences in eye movements or attention allocation to the stimuli. Second, the study showed that the responses in speech-motor regions were on the neurotypical level. This is important as it points towards a perceptual deficit that potentially contributes to one of the communication problems that individuals with ASD are faced with. Third, previous studies focused on human motion that was non-communicative (point light walkers) or emotional (facial expressions). Here, we show that the reduction in V5/MT and pSTS/STG responses (i.e. the TVSA) is present also for neutral facial movement that serves communicative function such as visual speech.

V5/MT and the pSTS/STG are proposed to build the dorsal pathway for processing facial movement (O'Toole et al., 2002; Bernstein and Yovel, 2015), and their functional connectivity is modulated by face movements (Furl et al., 2014). V5/MT responses to visual speech have been mainly found for relatively unspecific contrasts between dynamic videos and static images of faces (Calvert and Campbell, 2003; Callan et al., 2014) suggesting that it performs a general analysis of facial movement (O'Toole et al., 2002). Conversely, the pSTS/STG might process more complex information of the face (O'Toole et al., 2002; Ethofer et al., 2011). Previous studies showed that left pSTS/STG was more responsive to meaningful facial speech movement than to meaningless "gurning" and its response was correlated with behavioral performance in visual-speech recognizion tasks (Hall et al., 2005; Lee et al., 2007). In this context, the positive correlation between the right V5/MT responses and the visual-speech recognition performance in the ASD group, but not in the control group could be interpreted in two ways. First, it might be that the V5/MT is intact in ASD. If so, ASD individuals might recruit the right V5/MT to compensate for a potential dysfunction of the left TVSA that is more specialized to process visual-speech. Those participants, who could use this compensatory mechanism, would also show better visual-speech recognition performance. However, the overall reduced responses in the right V5/MT in the ASD group speak against such an interpretation. Another possibility is that

the V5/MT is dysfunctional in ASD. In this case, only those ASD participants who show higher V5/MT responses are still good at visual-speech recognition. This interpretation is in agreement with both the overall reduced responses in the V5/MT in the ASD group, and with the positive correlation between the behavioral visual-speech recognition performance and V5/MT responses in the ASD group. We speculate that ASD might be characterized by impairments in V5/MT, which then lead to subsequent response reductions in the pSTS/STG and/or reduced network connectivity to the speech-related left IFG. Such an interpretation is speculative, because functional connectivity analyses do not reveal the direction of information flow between the brain areas, and does not identify functional connections that are necessarily direct (Friston et al., 1997). However, the scenario would be in agreement with V5/MT structural alterations in ASD consistently reported in meta-analyses of VBM studies (Nickl-Jockschat et al., 2012; Deramus and Kana, 2015), while there is rather scarce evidence for structural alterations in the pSTS/STG (Cauda et al., 2014).

Some previous reports found intact V5/MT responses in ASD to human movement (Pelphrey et al., 2007; Koldewyn et al., 2011). However, on a closer look, in Pelphrey et al. (2007) it might be the inferior occipital gyrus and not the V5/MT that shows neurotypical responses and in the study by Koldewyn et al. (2011) the stimulus duration might have been too long (2 s) to detect motion deficits in ASD. Herrington et al. (2007) used a very similar design to Koldewyn et al. (2011), but shorter stimulus duration (1 s) and reported decreased V5/MT responses in ASD compared to controls. Detection of sensory deficits in ASD might require stimulus material with specific temporal features (Robertson et al., 2012; Van der Hallen et al., 2015; for review see Robertson and Baron-Cohen, 2017).

So far, only four behavioral studies monitored participant's eye movements during visual-speech recognition in ASD (Irwin et al., 2011, Saalasti et al., 2012; Irwin and Brancazio, 2014; Foxe et al., 2015). Interestingly, two studies reported visual-speech recognition deficits in ASD despite gaze patterns similar to neurotypical controls (Irwin et al., 2011; Foxe et al., 2015). Irwin et al. (2011) analyzed only trials where participants looked at the speaker's face and still visual-speech recognition was impaired in ASD compared to their controls. These findings are in line with our study where ASD individuals had difficulties recognizing speech information despite gaze behavior that was similar to the one of the controls in the visual-speech task. This supports the view that dysfunctional perception of facial movement and not altered gaze patterns are at core of visual-speech recognition deficits in adults with ASD.

Research in ASD often provides variable findings, probably due to the heterogeneous nature of this clinical condition, small study samples or non-hypothesis led research (e.g. Ioannidis, 2005; Button et al., 2013). In our study, we specifically selected a subgroup of adults diagnosed with high-functioning ASD to increase the homogeneity of ASD symptom characteristics. We accurately matched them pairwise to the neurotypical participants to also account for other variability sources of the behavior and brain responses. Our study was based on a hypothesis-driven approach and previous findings reporting the role of visual-movement regions and speech-related regions in visual-speech recognition (e.g., Blank and von Kriegstein, 2013), and behavioral visual-speech recognition deficits in ASD (e.g., Schelinski et al., 2014). We carefully corrected for the study hypotheses using a Holm-Bonferroni correction for multiple comparisons to avoid reporting false-positive findings.

5. Conclusions

The present study showed that the visual-speech recognition deficits in ASD were associated with a dysfunction already at the level of visual areas required for perception of motion signals, while other speech regions were intact. Impaired perception of visual-movement might contribute to deficits of speech acquisition and comprehension in face-to-face interactions. This emphasizes the importance of investigating sensory deficits to understand communication difficulties – a core feature of ASD.

Author contribution

K.B., S.S. and K.v.K. designed research, K.B. and S.S. performed research; K.B. and K.v.K. wrote the paper.

Conflict of interest

The authors declare no competing financial interests.

Funding

This work was funded by a Max Planck Research Group grant and an ERC-Consolidator Grant (SENSOCOM, 647051) to KVK and an Elsa-Neumann-Scholarship to KB.

Acknowledgements

We are grateful to the participants for taking part in our study. We acknowledge support by the Open Access Publication Funds of the SLUB/ TU Dresden

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.nicl.2018.09.019.

References

Alaerts, K., Woolley, D.G., Steyaert, J., Di Martino, A., Swinnen, S.P., Wenderoth, N., 2013. Underconnectivity of the superior temporal sulcus predicts emotion recognition deficits in autism. Soc. Cogn. Affect. Neurosci. 9, 1589–1600.
Alaerts, K., Swinnen, S.P., Wenderoth, N., 2017. Neural processing of biological motion in autism: an investigation of brain activity and effective connectivity. Sci. Rep. 7.
Allison, T., Puce, A., McCarthy, G., 2000. Social perception from visual cues: role of the STS region. Trends Cogn. Sci. 4, 267–278.
American Psychiatric Association, 2013. Diagnostic and Statistical Manual of Mental Disorders, 5th ed. American Psychiatric Association, Washington, DC.
Aschenberner, B., Weiss, C., 2005. Phoneme-Viseme Mapping for German Video-Realistic Audio-Visual-Speech-Synthesis.
Ashwood, K.L., Gillan, N., Horder, J., Hayward, H., Woodhouse, E., McEwen, F.S., Findon, J., Eklund, H., Spain, D., Wilson, C.E., Cadman, T., 2016. Predicting the diagnosis of autism in adults using the Autism-Spectrum Quotient (AQ) questionnaire. Psychol. Med. 46, 2595–2604.
von Aster, M., Neubauer, A., Horn, R., 2006. Wechsler Intelligenztest Für Erwachsene (WIE). Frankfurt/M, Harcourt Test Services.
Baron-Cohen, S., 1997. Mindblindness: An Essay on Autism and Theory of Mind. MIT press.
Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J., Clubley, E., 2001. Evidence from Asperger syndrome/high-functioning autism, males and females, scientists and mathematicians. J. Autism Dev. Disord. 31, 5–17 The autism-spectrum quotient (AQ).
Baum, S., Stevenson, R.A., Wallace, M.T., 2015. Behavioral, perceptual, and neural alterations in sensory and multisensory function in autism spectrum disorder. Prog. Neurobiol. 134, 140–160.
Beckers, G., Homberg, V., 1992. Cerebral visual motion blindness: transitory akinetopsia induced by transcranial magnetic stimulation of human area V5. Proc. R. Soc. Lond. B Biol. Sci. 249, 173–178.
Bernstein, M., Yovel, G., 2015. Two neural pathways of face processing: a critical evaluation of current models. Neurosci. Biobehav. Rev. 55, 536–546.
Bernstein, L.E., Jiang, J., Pantazis, D., Lu, Z.L., Joshi, A., 2011. Visual phonetic processing localized using speech and nonspeech face gestures in video and point-light displays. Hum. Brain Mapp. 32, 1660–1676.
Blake, R., Turner, L.M., Smoski, M.J., Pozdol, S.L., Stone, W.L., 2003. Visual recognition of biological motion is impaired in children with autism. Psychol. Sci. 14, 151–157.
Blank, H., von Kriegstein, K., 2013. Mechanisms of enhancing visual-speech recognition by prior auditory information. NeuroImage 65, 109–118.
Boddaert, N., Belin, P., Chabane, N., Poline, J.B., Barthélémy, C., Mouren-Simeoni, M.C., Brunelle, F., Samson, Y., Zilbovicius, M., 2003. Perception of complex sounds: abnormal pattern of cortical activation in autism. Am. J. Psychiatry 160, 2057–2060.
Bölte, S., Rühl, D., Schmötzer, G., Poustka, F., 2003. Diagnostisches Interview für Autismus – Revidiert (ADI-R). Verlag Hans Huber, Bern.
Brickenkamp, R., 2002. Test d2 - Aufmerksamkeits-Belastung-Test (d2). Hogrefe, Göttingen.

Brieber, S., Herpertz-Dahlmann, B., Fink, G.R., Kamp-Becker, I., Remschmidt, H., Konrad, K., 2010. Coherent motion processing in autism spectrum disorder (ASD) an fMRI study. Neuropsychologia 48, 1644–1651.
Button, K.S., Ioannidis, J.P., Mokrysz, C., Nosek, B.A., Flint, J., Robinson, E.S., Munafò, M.R., 2013. Power failure: why small sample size undermines the reliability of neuroscience. Nat. Rev. Neurosci. 14, 365.
Callan, D.E., Jones, J.A., Callan, A., 2014. Multisensory and modality specific processing of visual speech in different regions of the premotor cortex. Front. Psychol. 5.
Calvert, G.A., Campbell, R., 2003. Reading speech from still and moving faces: the neural substrates of visible speech. J. Cogn. Neurosci. 15, 57–70.
Calvert, G.A., Bullmore, E.T., Brammer, M.J., Campbell, R., Williams, S.C., McGuire, P.K., Woodruff, P.W.R., Iversen, S., David, A.S., 1997. Activation of auditory cortex during silent lipreading. Science 276, 593–596.
Campbell, R., MacSweeney, M., Surguladze, S., Calvert, G., McGuire, P., Suckling, J., Brammer, M.J., David, A.S., 2001. Cortical substrates for the perception of face actions: an fMRI study of the specificity of activation for seen speech and for meaningless lower-face acts (gurning). Cogn. Brain Res. 12, 233–243.
Cauda, F., Costa, T., Palermo, S., D'Agata, F., Diano, M., Bianco, F., Duca, S., Keller, R., 2014. Concordance of white matter and gray matter abnormalities in autism spectrum disorders: a voxel-based meta-analysis study. Hum. Brain Mapp. 35, 2073–2098.
Chevallier, C., Kohls, G., Troiani, V., Brodkin, E.S., Schultz, R.T., 2012. The social motivation theory of autism. Trends Cogn. Sci. 16, 231–239.
Chu, Y.H., Lin, F.H., Chou, Y.J., Tsai, K.W.K., Kuo, W.J., Jääskeläinen, I.P., 2013. Effective cerebral connectivity during silent speech reading revealed by functional magnetic resonance imaging. PLoS One 8, e80265.
DeRamus, T.P., Kana, R.K., 2015. Anatomical likelihood estimation meta-analysis of grey and white matter anomalies in autism spectrum disorders. Neuroimage Clin. 7, 525–536.
Desikan, R.S., Ségonne, F., Fischl, B., Quinn, B.T., Dickerson, B.C., Blacker, D., Buckner, R.L., Dale, A.M., Maguire, R.P., Hyman, B.T., Albert, M.S., Killiany, R.J., 2006. An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. NeuroImage 31, 968–980.
Eickhoff, S.B., Paus, T., Caspers, S., Grosbras, M.H., Evans, A.C., Zilles, K., Amunts, K., 2007. Assignment of functional activations to probabilistic cytoarchitectonic areas revisited. NeuroImage 36, 511–521.
Ethofer, T., Gschwind, M., Vuilleumier, P., 2011. Processing social aspects of human gaze: a combined fMRI-DTI study. NeuroImage 55, 411–419.
Foxe, J.J., Molholm, S., Del Bene, V.A., Frey, H.P., Russo, N.N., Blanco, D., Saint-Amour, D., Ross, L.A., 2015. Severe multisensory speech integration deficits in high-functioning school-aged children with autism spectrum disorder (ASD) and their resolution during early adolescence. Cereb. Cortex 25, 298–312.
Freitag, C.M., Konrad, C., Häberlen, M., Kleser, C., von Gontard, A., Reith, W., Troje, N.F., Krick, C., 2008. Perception of biological motion in autism spectrum disorders. Neuropsychologia 46, 1480–1494.
Friston, K.J., Buechel, C., Fink, G.R., Morris, J., Rolls, E., Dolan, R.J., 1997. Psychophysiological and modulatory interactions in neuroimaging. NeuroImage 6, 218–229.
Friston, K.J., Ashburner, A., Kiebel, S., Nichols, T., Penny, W. (Eds.), 2007. Statistical Parametric Mapping, The Analysis of Functional Brain Images. Academic Press.
Furl, N., Henson, R.N., Friston, K.J., Calder, A.J., 2014. Network interactions explain sensitivity to dynamic faces in the superior temporal sulcus. Cereb. Cortex 25, 2876–2882.
Giraud, A.L., Price, C.J., Graham, J.M., Truy, E., Frackowiak, R.S., 2001. Cross-modal plasticity underpins language recovery after cochlear implantation. Neuron 30, 657–664.
Grèzes, J., Fonlupt, P., Bertenthal, B., Delon-Martin, C., Segebarth, C., Decety, J., 2001. Does perception of biological motion rely on specific brain regions? NeuroImage 13, 775–785.
Grossman, E.D., Battelli, L., Pascual-Leone, A., 2005. Repetitive TMS over posterior STS disrupts perception of biological motion. Vis. Res. 45, 2847–2853.
Hall, D.A., Fussell, C., Summerfield, A.Q., 2005. Reading fluent speech from talking faces: typical brain networks and individual differences. J. Cogn. Neurosci. 17, 939–953.
Herrington, J.D., Baron-Cohen, S., Wheelwright, S.J., Singh, K.D., Bullmore, E.T., Brammer, M., Williams, S.C.R., 2007. The role of MT+/V5 during biological motion perception in Asperger Syndrome: an fMRI study. Res. Autism Spectr. Disord. 1, 14–27.
Holm, S., 1979. A simple Sequentially Rejective Multiple Test Procedure. Scand. J. Stat. 6, 65–70.
Ioannidis, J.P., 2005. Why most published research findings are false. PLoS Med. 2, e124.
Irwin, J.R., Brancazio, L., 2014. Seeing to hear? Patterns of gaze to speaking faces in children with autism spectrum disorders. Front. Psychol. 5.
Irwin, J.R., Tornatore, L.A., Brancazio, L., Whalen, D.H., 2011. Can children with autism spectrum disorders "hear" a speaking face? Child Dev. 82, 1397–1403.
Jezzard, P., Balaban, R.S., 1995. Correction for geometric distortion in echo planar images from B0 field variations. Magn. Reson. Med. 34, 65–73.
Jiang, J., Borowiak, K., Tudge, L., Otto, C., von Kriegstein, K., 2017. Neural mechanisms of eye contact when listening to another person talking. Soc. Cogn. Affect. Neurosci. 12, 319–328.
Koldewyn, K., Whitney, D., Rivera, S.M., 2011. Neural correlates of coherent and biological motion perception in autism. Dev. Sci. 14, 1075–1088.
von Kriegstein, K., Dogan, Ö., Grüter, M., Giraud, A.L., Kell, C.A., Grüter, T., Kleinschmidt, A., Kiebel, S.J., 2008. Simulation of talking faces in the human brain improves auditory speech recognition. Proc. Natl. Acad. Sci. 105, 6747–6752.
Lee, H.J., Truy, E., Mamou, G., Sappey-Marinier, D., Giraud, A.L., 2007. Visual speech circuits in profound acquired deafness: a possible role for latent multimodal connectivity. Brain 130, 2929–2941.
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. J. Autism Dev. Disord. 24, 659–685.
Lord, C., Risi, S., Lambrecht, L., Cook, E.H., Leventhal, B.L., DiLavore, P.C., Pickles, A., Rutter, M., 2000. The autism diagnostic observation schedule—Generic: a standard measure of social and communication deficits associated with the spectrum of autism. J. Autism Dev. Disord. 30, 205–223.
Malikovic, A., Amunts, K., Schleicher, A., Mohlberg, H., Eickhoff, S.B., Wilms, M., Palomero-Gallagher, N., Armstrong, E., Zilles, K., 2007. Cytoarchitectonic analysis of the human extrastriate cortex in the region of V5/MT+: a probabilistic, stereotaxic map of area hOc5. Cereb. Cortex 17, 562–574.
Marassa, L.K., Lansing, C.R., 1995. Visual word recognition in two facial motion conditions: full-face versus lips-plus-mandible. J. Speech Lang. Hear. Res. 38, 1387–1394.
Nichols, T., Hayasaka, S., 2003. Controlling the familywise error rate in functional neuroimaging: a comparative review. Stat. Methods Med. Res. 12, 419–446.
Nickl-Jockschat, T., Habel, U., Michel, T.J., Manning, J., Laird, A.R., Fox, P.T., Schneider, F., Eickhoff, S.B., 2012. Brain structure anomalies in autism spectrum disorder - a meta-analysis of VBM studies using anatomic likelihood estimation. Hum. Brain Mapp. 33, 1470–1489.
Nishitani, N., Hari, R., 2002. Viewing lip forms: cortical dynamics motor cortex, both during execution of hand actions. Neuron 36, 1211–1220.
Okada, K., Hickok, G., 2009. Two cortical mechanisms support the integration of visual and auditory speech: a hypothesis and preliminary data. Neurosci. Lett. 452, 219–223.
Oldfield, R.C., 1971. The assessment and analysis of handedness–he Edinburgh inventory. Neuropsychologia 9, 97–113.
O'Toole, A.J., Roark, D.A., Abdi, H., 2002. Recognizing moving faces: a psychological and neural synthesis. Trends Cogn. Sci. 6, 261–266.
Pelphrey, K.A., Morris, J.P., McCarthy, G., LaBar, K.S., 2007. Perception of dynamic changes in facial affect and identity in autism. Soc. Cogn. Affect. Neurosci. 2, 40–149.
Puce, A., Allison, T., Bentin, S., Gore, J.C., McCarthy, G., 1998. Temporal cortex activation in humans viewing eye and mouth movements. J. Neurosci. 18, 2188–2199.
Riedel, P., Ragert, P., Schelinski, S., Kiebel, S.J., von Kriegstein, K., 2015. Visual face-movement sensitive cortex is relevant for auditory-only speech recognition. Cortex 68, 86–99.
Robertson, C.E., Baron-Cohen, S., 2017. Sensory perception in autism. Nat. Rev. Neurosci. 18, 671–684.
Robertson, C.E., Martin, A., Baker, C.I., Baron-Cohen, S., 2012. Atypical integration of motion signals in autism spectrum conditions. PLoS One 7, e48173.
Robertson, C.E., Thomas, C., Kravitz, D.J., Wallace, G.L., Baron-Cohen, S., Martin, A., Baker, C.I., 2014. Global motion perception deficits in autism are reflected as early as primary visual cortex. Brain 137, 2588–2599.
Ross, L.A., Saint-Amour, D., Leavitt, V.M., Javitt, D.C., Foxe, J.J., 2007. Do you see what I am saying? Exploring visual enhancement of speech comprehension in noisy environments. Cereb. Cortex 17, 1147–1153.
Rouger, J., Lagleyre, S., Fraysse, B., Deneve, S., Deguine, O., Baron, P., 2007. Evidence that cochlear-implanted deaf patients are better multisensory integrators. Proc. Natl. Acad. Sci. 104, 7295–7300.
Rühl, D., Bölte, S., Feineis-Matthews, S., Poustka, F., 2004. Diagnostische Beobachtungsskala für Autistische Störungen (ADOS). Verlag Hans Huber, Bern.
Saalasti, S., Kätsyri, J., Tiippana, K., Laine-Hernandez, M., von Wendt, L., Sams, M., 2012. Audiovisual speech perception and eye gaze behavior of adults with Asperger syndrome. J. Autism Dev. Disord. 42, 1606–1615.
Sato, W., Toichi, M., Uono, S., Kochiyama, T., 2012. Impaired social brain network for processing dynamic facial expressions in autism spectrum disorders. BMC Neurosci. 13, 99.
Schelinski, S., Riedel, P., von Kriegstein, K., 2014. Visual abilities are important for auditory-only speech recognition: evidence from autism spectrum disorder. Neuropsychologia 65, 1–11.
Skipper, J.I., Nusbaum, H.C., Small, S.L., 2005. Listening to talking faces: motor cortical activation during speech perception. NeuroImage 25, 76–89.
Smith, E.G., Bennetto, L., 2007. Audiovisual speech integration and lipreading in autism. J. Child Psychol. Psychiatry 48, 813–821.
Smith, S.M., Jenkinson, M., Woolrich, M.W., Beckmann, C.F., Behrens, T.E., Johansen-Berg, H., Bannister, P.R., De Luca, M., Drobnjak, I., Flitney, D.E., Niazy, R.K., Sunders, J., Vickers, J., Zhang, Y., De Stefano, N., Brady, J.M., Matthews, P.M., 2004. Advances in functional and structural MR image analysis and implementation as FSL. NeuroImage 23, S208–S219.
Sumby, W.H., Pollack, I., 1954. Visual contribution to speech intelligibility in noise. J. Acoust. Soc. Am. 26, 212–215.
Travers, B.G., Adluru, N., Ennis, C., Tromp, D.P., Destiche, D., Doran, S., Bigler, E.D., Lange, N., Lainhart, J.E., Alexander, A.L., 2012. Diffusion tensor imaging in autism spectrum disorder: a review. Autism Res. 5, 289–313.
Tryfon, A., Foster, N.E., Sharda, M., Hyde, K.L., 2018. Speech perception in autism spectrum disorder: an activation likelihood estimation meta-analysis. Behav. Brain Res. 338, 118–127.
Van der Hallen, R., Evers, K., Brewaeys, K., Van den Noortgate, W., Wagemans, J., 2015. Global processing takes time: a meta-analysis on local–global visual processing in ASD. Psychol. Bull. 141, 549–573.

Van Wassenhove, V., Grant, K.W., Poeppel, D., 2005. Visual speech speeds up the neural processing of auditory speech. Proc. Natl. Acad. Sci. U. S. A. 102, 1181–1186.
Wakabayashi, A., Baron-Cohen, S., Wheelwright, S., Tojo, Y., 2006. The Autism-Spectrum Quotient (AQ) in Japan: a cross-cultural comparison. J. Autism Dev. Disord. 36, 263–270.
Watson, J.D., Myers, R., Frackowiak, R.S., Hajna, J.V., Woods, R.P., Mazziotta, J.C., Shipp, S., Zeki, S., 1993. Area V5 of the human brain: evidence from a combined study using positron emission tomography and magnetic resonance imaging. Cereb. Cortex 3, 79–94.
Wechsler, D., 1997. Wechsler Adult Intelligence Scale (WAIS-III). The Psychological Corporation, San Antonio, TX.
Williams, J.H.G., Massaro, D.W., Peel, N.J., Bosseler, A., Suddendorf, T., 2004. Visual-auditory integration during speech imitation in autism. Res. Dev. Disabil. 25, 559–575.
Wilson, S.M., Saygin, A.P., Sereno, M.I., Iacoboni, M., 2004. Listening to speech activates motor areas involved in speech production. Nat. Neurosci. 7, 701.
World Health Organization, 2004. International Statistical Classification of Diseases and Related and Related Health Problems, 10th ed. World Health Organization, Geneva.
Zeki, S., Watson, J.D., Lueck, C.J., Friston, K.J., Kennard, C., Frackowiak, R.S., 1991. A direct demonstration of functional specialization in human visual cortex. J. Neurosci. 11, 641–649.