RESEARCH ARTICLE

Dorsal-movement and ventral-form regions are functionally connected during visual-speech recognition

Kamila Borowiak[1,2,3] | Corrina Maguinness[1,2] | Katharina von Kriegstein[1,2]

[1]Chair of Cognitive and Clinical Neuroscience, Faculty of Psychology, Technische Universität Dresden, Dresden, Germany

[2]Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany

[3]Berlin School of Mind and Brain, Humboldt University of Berlin, Berlin, Germany

Correspondence
Kamila Borowiak, Technische Universität Dresden, Bamberger Straße 7, 01187 Dresden, Germany.
Email: kamila.borowiak@tu-dresden.de

Funding information
Elsa-Neumann-Scholarship; ERC-Consolidator grant, Grant/Award Number: 647051; Max Planck Research Group grant

Abstract

Faces convey social information such as emotion and speech. Facial emotion processing is supported via interactions between dorsal-movement and ventral-form visual cortex regions. Here, we explored, for the first time, whether similar dorsal–ventral interactions (assessed via functional connectivity), might also exist for visual-speech processing. We then examined whether altered dorsal–ventral connectivity is observed in adults with high-functioning autism spectrum disorder (ASD), a disorder associated with impaired visual-speech recognition. We acquired functional magnetic resonance imaging (fMRI) data with concurrent eye tracking in pairwise matched control and ASD participants. In both groups, dorsal-movement regions in the visual motion area 5 (V5/MT) and the temporal visual speech area (TVSA) were functionally connected to ventral-form regions (i.e., the occipital face area [OFA] and the fusiform face area [FFA]) during the recognition of visual speech, in contrast to the recognition of face identity. Notably, parts of this functional connectivity were decreased in the ASD group compared to the controls (i.e., right V5/MT—right OFA, left TVSA—left FFA). The results confirmed our hypothesis that functional connectivity between dorsal-movement and ventral-form regions exists during visual-speech processing. Its partial dysfunction in ASD might contribute to difficulties in the recognition of dynamic face information relevant for successful face-to-face communication.

KEYWORDS

atypical perception, dynamic face perception, fMRI, form, functional connectivity, high-functioning ASD, lip reading, movement

1 | INTRODUCTION

Faces represent an essential source of information that is relevant for human communication (Bruce & Young, 1986). In everyday face-to-face communication, faces are dynamic by nature, for example, due to the fast articulatory movements associated with speech. The fast and accurate perception of these visible movements is often crucial for successful communication (Ross, Saint-Amour, Leavitt, Javitt, & Foxe, 2007; Sumby & Pollack, 1954).

Traditional neuroscience models proposed that dynamic faces convey two types of information: (a) variant, that is, dynamic information (such as emotional expression and visual speech) and (b) invariant information (such as identity). These are processed in two distinct brain pathways: a dorsal pathway for dynamic information and a ventral pathway for invariant information (Haxby, Hoffman, & Gobbini, 2000; OToole, Roark, & Abdi, 2002). Several findings, however, suggest that dorsal pathway regions might also be critical for processing invariant face information (e.g., Anzellotti & Caramazza, 2017; Dobs,

Schultz, Bülthoff, & Gardner, 2018; Fox, Hanif, Iaria, Duchaine, & Barton, 2011), and vice versa ventral pathway regions might also be critical for processing dynamic information (e.g., LaBar, Crupain, Voyvodic, & McCarthy, 2003; Sato, Kochiyama, Yoshikawa, Naito, & Matsumura, 2004; Schultz & Pilz, 2009). For example, responses in the posterior superior temporal sulcus/gyrus in the dorsal pathway have been observed for the processing of invariant face-identity (Dobs et al., 2018; Fox et al., 2011), while responses in the fusiform face area in the ventral pathway have been observed during the processing of dynamic, in contrast to static, emotional facial expressions (e.g., LaBar et al., 2003; Schultz & Pilz, 2009). Therefore, a recent model proposes a division between the pathways at a different functional level (Bernstein & Yovel, 2015). Bernstein and Yovel (2015) assumed that while both pathways are concurrently recruited during dynamic face processing, each pathway might extract discrete information from the dynamic face that is relevant for its successful perception. Similar to the traditional models, the dorsal pathway is proposed to process movement information from the face, while the ventral pathway processes structural form information. This dorsal pathway is sensitive to different aspects of facial movements. These movement profiles include speed (e.g., speed with which a person reaches the peak of a smile) and trajectory, that is, the course of variations in facial features that are induced by the movement over time (e.g., change of the lip-corner position during a smile; Knappmeyer, Thornton, & Bülthoff, 2003). The ventral pathway is sensitive to structural form information. This includes the global shape of the face or individual facial features, and their variations induced by the movement (e.g., the modified shape of the lip corners due to a smile). In contrast to the traditional models, Bernstein and Yovel's revised model highlights that the ventral pathway can also extract static structural form cues (represented as "static snapshots") from the dynamic image that are relevant also for emotion and speech processing. Here, we will refer to these pathways as the "dorsal-movement pathway" and the "ventral-form pathway," respectively.

The dorsal-movement pathway includes the extrastriate visual area V5/MT and a region in the posterior superior temporal sulcus/gyrus (pSTS/STG; Beckers & Homberg, 1992; Grossman, Battelli, & Pascual-Leone, 2005). The V5/MT is relevant for general movement perception (Zeki et al., 1991), while the pSTG/STS is more selective and sensitive to human movement only (Allison, Puce, & McCarthy, 2000; Grossman et al., 2000). The portion of the pSTS/STG that is specifically sensitive to visual speech has been coined the temporal visual speech area (TVSA; Bernstein, Jiang, Pantazis, Lu, & Joshi, 2011). The V5/MT and the TVSA are more responsive to dynamic compared to static faces (Kilts, Egan, Gideon, Ely, & Hoffman, 2003; Pitcher, Duchaine, & Walsh, 2014; Schultz & Pilz, 2009), and their functional connectivity to each other is modulated by the perception of facial movements (Borowiak, Schelinski, & von Kriegstein, 2018; Foley, Rippon, Thai, Longe, & Senior, 2012; Furl, Henson, Friston, & Calder, 2014). The ventral-form pathway includes the occipital face area (OFA) and the fusiform face area (FFA) which refers to face-sensitive portions of the inferior occipital gyrus and the fusiform gyrus, respectively (Gauthier, Skudlarski, Gore, & Anderson, 2000;

Kanwisher, McDermott, & Chun, 1997). The OFA is involved in perceptual analysis of facial structure and preferentially represents individual facial features, including the eyes and the nose (Liu, Harris, & Kanwisher, 2002; Liu, Harris, & Kanwisher, 2010; Pitcher, Walsh, Yovel, & Duchaine, 2007). According to some views, the OFA acts as the first face-selective cortical region before information reaches the FFA (Haxby et al., 2000; Pitcher, Walsh, & Duchaine, 2011; but see also Rossion, 2008). The FFA might be involved in more intricate computations, such as the integration of individual facial features (i.e., holistic processing) for identity recognition (Andrews & Ewbank, 2004; Farah, Wilson, Drain, & Tanaka, 1998; Grill-Spector, Knouf, & Kanwisher, 2004; Harris & Aguirre, 2010). This region might also integrate structural cues relevant for emotion recognition (Ganel, Valyear, Goshen-Gottstein, & Goodale, 2005; Vuilleumier, Armony, Driver, & Dolan, 2001). The OFA and the FFA are functionally and anatomically connected in the typically developing brain, and as such are well suited to construct a single pathway for face recognition (Avidan & Behrmann, 2014; Ethofer, Gschwind, & Vuilleumier, 2011; Fairhall & Ishai, 2006; Gschwind, Pourtois, Schwartz, Van De Ville, & Vuilleumier, 2012; Pyles, Verstynen, Schneider, & Tarr, 2013).

Although evidence for the involvement of both pathways exists, our knowledge on how, or if, these pathways interact during dynamic face processing is limited. Two fMRI studies provided empirical evidence that dorsal-movement (V5/MT, pSTS/G [i.e., TVSA]) and ventral-form regions (OFA) are functionally connected to each other during facial emotion perception (Foley et al., 2012; Furl et al., 2014). To date, it remains unknown whether such a functional connectivity between the regions in the dorsal-movement and in the ventral-form pathway might also be present for other kinds of facial movements such as visual speech. Findings that visual-speech information can be extracted from dynamic point-light displays, in the absence of facial form information (Rosenblum, Johnson, & Saldaña, 1996; Rosenblum & Saldaña, 1996), suggest that visual-speech recognition does not necessarily rely on information processed in the ventral-form regions. However, it is possible to identify vowels or consonants based on facial form information only (Campbell, 1996a; Campbell, Landis, & Regard, 1986; Schweinberger & Soukup, 1998). This suggests that although form information might not be essential for visual-speech recognition, it is nevertheless informative (Rosenblum & Saldaña, 1998; Thomas & Jordan, 2002). Functional connectivity between dorsal-movement and ventral-form regions could provide a supporting mechanism for visual-speech recognition via the extraction of additional informative static face cues that are relevant for speech perception.

Here, we investigated functional connectivity between dorsal-movement and ventral-form regions during visual-speech recognition. Our first aim was to test the hypothesis that dorsal–ventral functional connectivity exists in typically developing individuals during visual-speech recognition and that the strength of the functional connectivity correlates positively with visual-speech recognition ability.

Revealing mechanisms behind visual-speech recognition is important because face-to-face communication relies on both auditory-speech and visual-speech signals (Arnold & Hill, 2001; Ross et al.,

2007; Sumby & Pollack, 1954). Successful perception of visual speech can substantially enhance our understanding of auditory speech (Ross et al., 2007; Sumby & Pollack, 1954; van Wassenhove, Grant, & Poeppel, 2005; von Kriegstein et al., 2008). This can be particularly beneficial in situations with high background noise (MacLeod & Summerfield, 1987) or for populations with hearing impairments (Giraud, Price, Graham, Truy, & Frackowiak, 2001; Maguinness, Setti, Burke, Kenny, & Newell, 2011; Rouger et al., 2007).

Autism spectrum disorder (ASD) is characterized by communication and social interaction difficulties (DSM-5, American Psychiatric Association, 2013). The cause for these difficulties is still unknown. One contributing factor could be deficits in processing social signals such as dynamic faces. Such difficulties have been reported both for facial emotion (Sato, Toichi, Uono, & Kochiyama, 2012; Sato, Uono, & Toichi, 2013), and visual-speech information (Foxe et al., 2015; Schelinski, Riedel, & von Kriegstein, 2014). These deficits are associated with reduced brain responses in ASD in visual sensory cortices (i.e., V5/MT, pSTS/STG [TVSA] and FFA; Borowiak et al., 2018; Pelphrey, Morris, McCarthy, & LaBar, 2007; Sato et al., 2012; but see Kliemann et al., 2018), and also with decreased functional connectivity between the two dorsal-movement pathway regions (Borowiak et al., 2018). Given the importance of visual-speech perception for successful face-to-face communication (Arnold & Hill, 2001; Ross et al., 2007; Sumby & Pollack, 1954), less efficient processing of visual speech may contribute to speech comprehension difficulties in face-to-face situations in ASD (Smith and Bennetto, 2007; Schelinski et al., 2014). In conjunction with other factors (e.g., atypical audio–visual integration [Stevenson et al., 2014]), atypical visual-speech processing may be one aspect to consider for better understanding of the socio-communication deficits observed in ASD.

Our second aim was to explore whether alterations in dorsal–ventral functional connectivity can be observed in individuals with ASD and whether this might contribute to their difficulties with visual-speech recognition. We had two alternative hypotheses. First, we predicted that dorsal–ventral functional connectivity could be intact in ASD and similar to the typically developing population. Second, dorsal–ventral functional connectivity in ASD could be reduced, in comparison to the typically developing population, indicating that it is dysfunctional in ASD. In both cases, a positive correlation between the degree of dorsal–ventral functional connectivity in ASD and visual-speech recognition performance would indicate that dorsal–ventral functional connectivity might be beneficial in ASD, and function as a compensatory mechanism for dysfunctional processing in dorsal-movement regions (Borowiak et al., 2018).

We conducted a functional magnetic resonance imaging (fMRI) experiment on visual-speech recognition and an fMRI region of interest (ROI) localizer in a group of typically developing adults who were pairwise matched to a group of adults with high-functioning ASD. In the fMRI visual-speech recognition experiment, participants saw silent videos of speakers articulating syllables and performed a visual–speech recognition task and a face-identity recognition task. The two tasks were performed on identical stimulus material. In the fMRI ROI localizer, we used previously published procedures to functionally localize ROIs in the dorsal-movement and in the ventral-form pathways (von Kriegstein et al., 2008).

One obvious prerequisite for successful visual-speech recognition is that participants look at informative parts of the face (Marassa & Lansing, 1995). Some studies reported that individuals with ASD gaze less to the face and the mouth during visual-speech recognition compared to typically developing controls (Irwin & Brancazio, 2014; Irwin, Tornatore, Brancazio, & Whalen, 2011, but see Foxe et al., 2015; Saalasti et al., 2012). Since gaze behavior influences brain responses to faces (Dalton et al., 2005; Jiang, Borowiak, Tudge, Otto, & von Kriegstein, 2017), we used an eye tracker in the MRI environment to assess where participants looked during visual-speech recognition.

2 | METHODS AND MATERIALS

The results of the present study are based on data that was also used in Borowiak et al. (2018) to address a different research question.

2.1 | Participants and neuropsychological assessment

The study sample included 17 typically developing individuals (control group) and 17 individuals diagnosed with ASD (ASD group). The groups were matched pairwise on gender, chronological age, handedness (Oldfield, 1971), and full-scale intelligence quotient (IQ; Table 1). The ASD compared to the control group had a significantly lower performance IQ ($p < .045$). We excluded three additional participants with ASD: one participant due to difficulties in finding a control participant 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 fMRI visual-speech recognition experiment that was lower than two SDs of the mean performance of the ASD group. Data of the respective control participants were 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 SD (15 IQ points). IQ was assessed using the Wechsler Adult Intelligence Scale (WAIS-III; Wechsler, 1997; German adapted version: Von Aster et al., 2006). IQ is important to consider when studying aspects of face processing, because it has been linked to recognition of emotion and identity from the face, at least in typically developing individuals (Lawrence et al., 2008; Lawrence, Campbell, & Skuse, 2015; but see Davis et al., 2011; Zhu et al., 2010). In addition, both groups showed comparable concentration performances (d2 test of attention; Brickenkamp, 2002; Table 1). All participants reported normal or corrected-to-normal vision. 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, 3,000, 4000, 6000, and 8000 Hz; Micromate 304;

Gender Handedness[a]	Control (n = 17) 13 males, 4 females 14 right, 3 left		ASD (n = 17) 13 males, 4 females 14 right, 3 left		
	M	SD (range)	M	SD (range)	p
Age	32.65	11.08 (21–55)	31.47	10.82 (21–54)	.756
WAIS-III[b] scales					
Full scale IQ	107.12	8.17 (91–121)	105.35	10.64 (87–124)	.591
Verbal IQ	106.29	10.84 (89–130)	109.06	12.61 (91–138)	.498
Performance IQ	106.76	8.78 (90–121)	100.12	9.76 (82–120)	.045*
Working memory	103.76	11.44 (88–126)	105.65	13.32 (86–146)	.662
Attention (d2)[c]	105.12	7.66 (86–114)	101.82	11.73 (84–126)	.341
AQ[d]	17.06	4.07 (10–25)	37.94	7.82 (14–47)	.000*

TABLE 1 Descriptive statistics for the control group and the ASD group

[a]Handedness was assessed using the Edinburgh handedness questionnaire (Oldfield, 1971).
[b]WAIS-III, Wechsler Adult Intelligence Scale (Wechsler, 1997; German adapted version: Von Aster, Neubauer, & Horn, 2006; M = 100; SD = 15).
[c]Concentration = d2 test of attention (Brickenkamp, 2002; M = 100; SD = 10).
[d]AQ, Autism Spectrum Quotient (Baron-Cohen, Wheelwright, Skinner, Martin, & Clubley, 2001).
*Significant group differences (p < .05); M = mean; SD = standard deviation.

Madsen, Denmark). All participants were native German speakers and were free of psychostimulant medication.

Participants with ASD had previously received a formal clinical diagnosis of Asperger Syndrome (13 male, 4 female) or childhood autism (one 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, Bölte, Feineis-Matthews, & Poustka, 2004), that was conducted in the context of clinical diagnostics or by researchers with formal training on conducting the ADOS (KB, SS). If caregivers or relatives were available (n = 11), we also performed the Autism Diagnostic Interview-Revised (ADI-R; Lord, Rutter, & Le Couteur, 1994; German version: Bölte, Rühl, Schmötzer, & Poustka, 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 reached a clinically relevant extent 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 typically developing 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 typically developing controls (e.g., Baron-Cohen et al., 2001; Wakabayashi, Baron-Cohen, Wheelwright, & Tojo, 2006), but it does not significantly predict a positive ASD diagnosis (Ashwood et al., 2016). None of the control participants reported a history of psychiatric disorders or a family history of

ASD. None of the participants reported a 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 of Leipzig (316–15-24082015). All participants received expense reimbursement (8€/hr for MRI session, 7 €/hr for behavioral session and travel cost reimbursement).

2.2 | Experiments

The experimental procedure consisted of one fMRI session including a fMRI visual-speech recognition experiment and an fMRI ROI localizer, and one behavioral session including behavioral tests conducted on a computer outside the MRI environment. Participants, who had never participated in a MRI investigation before were familiarized with the MRI environment by means of a mock MRI scanner session. This was done on a day before the MRI data acquisition.

2.2.1 | Visual-speech recognition experiment

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 three 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 & Weiss, 2005). In each block, the participants either performed a visual-speech task or a face-identity task (Figure 1a). Before each block (Figure 1b),

participants received a task instruction. They saw a written instruction screen "Silbe" (English "syllable") or "Person" (English "person") to indicate which task to perform. The screen was followed by the presentation of a video of one of the three 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 following 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 identity of the target person or not, independent of the syllable that was articulated. The response could be given until the end of the video. After each block, a white fixation cross on a black screen was presented for a period of 18 s. The stimulus material was identical 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


FIGURE 1   Experimental designs of the visual-speech recognition experiment and the ROI localizer. (a) Visual-speech recognition experiment: Participants viewed blocks of videos without an audio-stream showing three 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 (the German words for "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. There were four conditions, that is, static faces, facial speech movement, static object, and object movement [Color figure can be viewed at wileyonlinelibrary.com]
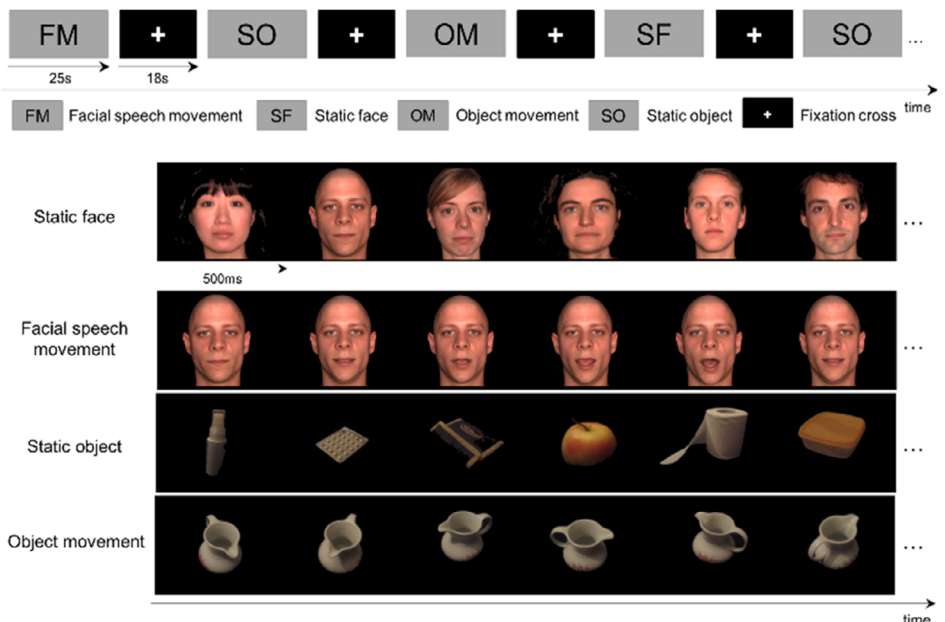
two and five across blocks and was the same for the visual-speech task and the face-identity task. Responses were made via a button box. Participants were requested to respond to each item by pressing one button if it was a target and another button if it was not. The experiment was divided into two 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 three practice blocks per task, which had the same structure as blocks for the fMRI experiment, but different stimulus material (three speakers and nine VCV-syllables not included in the fMRI experiment).

All videos started and ended with the closed mouth of the speaker. Videos were on average 2.18 s (±0.12 s) long. Syllables were recorded from six 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 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). 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 gray scale and AVI 4:3 format (1,024 × 768 pixels).

2.2.2 | ROI localizer (fMRI)

The ROI localizer was a $2 \times 2 \times 2$ 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 in which still frames of videos were shown (Figure 1c): (i) static face (faces of different persons with different articulatory positions), (ii) facial speech movement (faces of the same person with different articulatory positions), (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 one 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

Assessment of face recognition abilities included two standard tests for facial form perception and facial memory respectively
Cambridge Face Perception Test (*CFPT*; Duchaine, Germine, & Nakayama, 2007): The CFPT includes static pictures of neutral faces. It tests face perception and requires participants to order a series of faces

("comparison faces") to a target face according to similarity. The comparison faces comprise the target face morphed toward several different faces by varying degree. One half of the trials includes upright faces and the other half presents inverted faces. Scores for each item are computed by summing the deviations from the correct position (number of errors) for each face. A score of 93.3 (for upright or inverted trials) signifies chance performance as it reflects the original pseudorandomized ordering (i.e., deviation from the correct rank order) of the stimuli. Scores were calculated separately for upright and inverted trials.

Cambridge Face Memory Test (*CFMT*; Duchaine & Nakayama, 2006): The CFMT includes static pictures of neutral faces. It tests the ability to learn and recognize the identity of an unfamiliar person by face. Participants are introduced to six target faces, which are presented in three different views. Then, facial memory is tested with forced choice items consisting of three faces, one of which is a target and two are unfamiliar faces. Participants indicate which face they had seen before. A score of 33% indicates chance performance.

2.3 | Eye tracking

During MRI scanning, we recorded participants' eye movements using a 120 Hz monocular MR compatible eye tracker (EyeTrac 6, ASL). The optical path was reflected over a mirror placed on top of the head coil 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 and if necessary, the eye tracking system was recalibrated.

2.4 | Image acquisition

Functional and structural data was acquired on a 3-Tesla SIEMENS MAGNETOM Prisma MRI machine (Siemens Healthineers, Germany). Functional images were collected with a 20-channel head coil using a gradient echo EPI (echo planar imaging) sequence (TR = 2.790 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). To correct for field distortions, field-map scans were acquired which consisted of a pair of two-dimensional gradient echo images with different echo times (TE1/TE2 = 4.92 ms/ 7.38 ms; Jezzard & Balaban, 1995).

Following the functional images, a structural image was acquired using a 32-channel head coil and a T1-weighted three-dimensional magnetization-prepared rapid gradient echo (MPRAGE) sequence (TR = 2.300 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 those participants (n = 10) who had never undergone any MRI investigation at the Max Planck Institute for Human Cognitive and Brain Sciences before. For all other participants, we accessed MPRAGE images available in the institute's data bank, which had been also acquired with a 32-channel coil and with the exact same acquisition parameters on 3-Tesla MRI machines (SIEMENS MAGNETOM Trio, Verio and Prisma; Siemens Healthineers, Germany).

2.5 | Data analysis

2.5.1 | Behavioral data

Behavioral data were analyzed with PASW Statistics 22.0 (IBM SPSS Statistics). Data were tested for normality and homogeneity of variance using the Shapiro–Wilk test and the Levene's test, respectively. Except for the performance of the control group in the face-identity task in the visual-speech recognition fMRI experiment ($p = .002$), all the behavioral data followed a normal distribution ($p > .05$). Data variance between the groups was equal for the performance in the CFMT and in the CFPT ($p > .223$), but not in the visual-speech task and in the face-identity task in the visual-speech recognition fMRI experiment ($p \leq .004$).

We computed group comparisons using analysis of variance (ANOVA) and Welch's independent-samples t-test. Welch's t-test is more reliable when two samples have unequal variances (Ruxton, 2006). All t-tests were calculated two-tailed. Linear regression was used to test whether the visual-speech recognition performance was predicted by any measure of face recognition abilities. Level of significance for all tests was defined at $\alpha = .05$. To estimate the effect sizes, we used η^2 (Eta squared) and Cohen's d.

2.5.2 | Eye tracking

Eye tracking data were analyzed offline (ASL Results Plus, Applied Science Laboratories, Bedford). Data from 12 ASD participants and 12 control participants were 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 (four ASD and one control participant). Eye tracking data from their respective matched participants were 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 fMRI experiment (visual-speech task, face-identity task), and for the face conditions of the fMRI ROI localizer (static faces, facial speech movement). To investigate where participants looked, we created rectangular areas of interest (AOIs). For the visual-speech recognition experiment and the ROI localizer, we defined three AOIs: "Eye," "Mouth," and "Off." We compared the total number of fixations and the number of fixations within each AOI between the groups and between the conditions using a repeated measures ANOVA. The methods of the eye tracking analysis were previously described in more detail in Borowiak et al. (2018).

2.5.3 | Functional MRI

Preprocessing and movement artifact correction
MRI data were 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). 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.

To control for potential confounding effects of movement artifacts on the Blood Oxygenation Level Dependent (BOLD) signal change, we examined the head movement along six possible axes during both experiments. We compared six movement parameters resulting from rigid body transformation during spatial realignment using independent-samples t-tests. 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 (Table S1). Such finding is in accordance with previous literature (for a review, see Travers et al., 2012). To control for this movement differences between the groups, we examined each participant's functional time series for global-signal artifacts 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 SDs from the overall mean for all images (ART z-threshold = 3.0), and the absolute global translation movement was more than 3 mm. There were no significant group differences in the number of outlier volumes in any of the two experiments (Table S1). Outlier volumes and six movement parameters were modeled as covariates of no interest in the first-level general linear model (GLM).

Local 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, Ashburner, Kiebel, Nichols, & Penny, 2007). For the ROI localizer, we modeled the conditions "static faces," "facial speech movement," "static objects," and "object movement." We computed three contrasts of interest: "(facial speech movement + object movement) > (static faces + static objects)," "facial speech movement > static faces," "faces > objects." 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." We computed one contrast of interest "visual-speech task > face-identity task" for the purpose of seed region definition in the psychophysiological interaction (PPI) analysis in each individual participant. Head movement parameters and outlier volumes were modeled as covariates of no interest. In addition, the normalized numbers of eye fixations onto the predefined AOIs ("Eye," "Mouth," "Off") in the visual-speech task and in the face-identity task were also included 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 signal changes were based on a random effects model that estimated the

second-level statistic at each voxel. For the ROI localizer, we performed one-sample *t*-tests across the single-subject contrast as within-subject analyses to define the ROIs for a second-level ROI analysis of the functional connectivity. For the visual-speech recognition experiment, we did not assess local BOLD responses at the second level.

ROI definition

We defined individual-level ROIs in the dorsal-movement regions for the purpose of seed region definition in the PPI analyses. Group-level ROIs in the ventral-form regions were used as target regions for the PPI analyses.

Individual ROIs in the dorsal-movement regions

To define individual ROIs in the dorsal-movement regions, we used the following procedure: for each participant, we identified four dorsal-movement regions (i.e., right V5/MT, left V5/MT, right TVSA and left TVSA). The bilateral V5/MT was localized based on the contrast "(facial speech movement + object movement) > (static faces + static objects)," because the V5/MT region is known to be involved in perception of general movement (Zeki et al., 1991). The bilateral TVSA was localized using the contrast "facial speech movement > static faces" because it is known to be relevant for perception of visual-speech movements (Bernstein et al., 2011; Puce, Allison, Bentin, Gore, & McCarthy, 1998). Please note that the initial TVSA definition was based on a different contrast between speech and nonspeech facial movements (visual speech > visual nonspeech) ∩ (point-light speech > point-light nonspeech; Bernstein et al., 2011). Here, we adopted the term "TVSA" 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. The TVSA definition in our study might also contain other regions compared to the TVSA by Bernstein et al. (2011), because our control "static face" condition included only the face, but no movement.

To ensure that the individual ROIs were located within the anatomically defined V5/MT and TVSA, responses to the contrasts of interest were overlaid with probabilistic anatomical masks of the respective brain regions implemented in FSL (V5/MT: Jülich histological [cyto- and myelo-architectonic] atlas [Eickhoff et al., 2007]; the pSTS/STG for TVSA: Harvard–Oxford cortical structural atlas [Desikan et al., 2006]). For both anatomical maps, we chose a threshold of 10% to restrict maps to anatomically meaningful structures.

The individual dorsal-movement ROIs were defined in each participant as 4-mm-radius spheres centered on the peak responses of the respective contrasts of interest that were located within the probabilistic anatomical masks (V5/MT: Table S2; TVSA: Table S3). If there was no peak in the individual participant even at a lenient threshold ($p < .09$ uncorrected to reduce Type II error, that is, missing an individual participant's peak), we used the group coordinate of the contrast of interest. To ensure that the individual spheres contained only regions anatomically defined as the V5/MT and the TVSA, the 4-mm-radius spheres were additionally overlaid with the respective probabilistic anatomical masks implemented in FSL. The overlap between the

spheres and the anatomical maps was defined as the individual dorsal-movement ROIs.

Group-level ROIs in the ventral-form regions

The ROIs in the bilateral FFA were defined by means of a combined functional and anatomical approach in the control group, in the ASD group and in both groups together using FSL (Smith et al., 2004; http://www.fmrib.ox.ac.uk/fsl/fslview). First, we extracted probabilistic functional maps of the right and the left FFA included in the probabilistic atlas of face-sensitive brain regions (Engell & McCarthy, 2013; threshold of 20%). Second, we defined peak coordinates of the BOLD responses to the contrast "faces > objects" in the ROI localizer that were located within the probabilistic maps. The ROI localizer was thresholded to reach a ROI size of approximately 25 voxels. The overlap between the functional BOLD responses and the probabilistic maps was defined as the right FFA ROI and the left FFA ROI, respectively.

In the first step, we created the right and the left FFA ROIs based on the BOLD responses of all the participants in the control group, in the ASD group and in both groups together ("first-step group" ROIs; Figure S1a). However, with this approach, we were not able to create a left FFA ROI in the control group of a size comparable to the other FFA ROIs even at a very lenient threshold (16 voxels at $p < .9$ uncorrected). This was not surprising, because the left FFA is known to be more difficult to localize and to be on average smaller than the right FFA (Rossion, Hanseeuw, & Dricot, 2012; Yovel, Tambini, & Brandman, 2008). In a second step, we defined the FFA ROIs based on BOLD responses of only those participants who showed BOLD responses in the respective brain region on the single-participant level at a threshold of $p < .09$ uncorrected (Table S4). The relatively lenient statistical threshold necessary to obtain the responses was expected as the experiment was a short localizer. All the individual coordinates were in agreement with previously published FFA coordinates (Blank, Wieland, & von Kriegstein, 2014; Sabatinelli et al., 2011). We will call these ROIs "second-step group" ROI in the following. The number of included participants per group varied between the regions (right FFA: $n_{CON} = 14$ and $n_{ASD} = 13$; left FFA: $n_{CON} = 12$ and $n_{ASD} = 13$). With this approach, we were able to define FFA ROIs of comparable sizes in both hemispheres and groups (controls and ASD: right FFA = 26 voxels; left FFA = 25 voxels; Figure S1b). We used the "second-step group" ROIs as FFA ROIs for the second-level ROI analysis of the functional connectivity in the visual-speech recognition experiment. In addition, we conducted at the second level a control ROI analysis using "first-step group" ROIs to ensure that the reported effects were robust to different ROI definitions (see Supporting Information Results).

The bilateral OFA was defined using FSL (Version 5.0.8, FMRIB, Oxford, UK, http://fsl.fmrib.ox.ac.uk/fsl/fslwiki/). First, we extracted the probabilistic anatomical maps of the bilateral occipital fusiform gyrus from the Harvard–Oxford cortical structural atlas that corresponds to the inferior occipital gyrus (Desikan et al., 2006). Second, we intersected the anatomical maps with probabilistic functional maps from the probabilistic atlas of face-sensitive brain regions (Engell &

McCarthy, 2013). The left OFA is known to be on average smaller than the right OFA (Yovel et al., 2008). To make the size of the right and the left OFA ROIs comparable, we restricted the probabilistic functional maps using different thresholds: 25% for the right OFA and 20% for the left OFA. The overlap between the anatomical and the functional probabilistic maps were defined as the right OFA ROI (37 voxels) and the left OFA ROI (36 voxels).

Psycho-physiological interaction analyses

We investigated functional connectivity (i.e., temporal correlations between spatially remote neurophysiological events [Friston, 1994]) during the visual-speech task compared to the face-identity task between regions of the dorsal-movement pathway and regions of the ventral-form pathway. Functional connectivity was assessed using PPI analysis (Friston et al., 1997). PPI analysis is a method for investigating changes in the relationship between responses in different brain regions that occur due to a specific task. It is based on a correlation analysis of BOLD responses. It identifies which correlations between responses in one specific brain region (i.e., seed region) and responses in other brain regions (i.e., target regions) are modulated by a psychological factor (i.e., experimental task). Thus, it assesses if responses in two brain regions increase and decrease "in synchrony" under the influence of a specific task, independent of the amplitude of the time series (O'Reilly, Woolrich, Behrens, Smith, & Johansen-Berg, 2012). This means that low responses in a certain brain region are not necessarily associated also with a low functional connectivity to other brain regions. Although a PPI effect is reflective of a task-specific increase in the flow of information between brain regions (O'Reilly et al., 2012), it does not necessarily correspond to anatomical connections between said regions.

Seed regions from which we extracted the physiological variable were defined in the right and in the left V5/MT and in the right and in the left TVSA. The four 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 dorsal-movement ROI as defined based on the ROI localizer (see Section "Individual ROIs in dorsal-movement regions").

Covariates (first Eigenvariate from seed region, psychological variable, PPI term) were created using routines implemented in SPM12. 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 term, the psychological variable and the first Eigenvariate were entered as covariates in a design matrix. To control for significant group differences in head movements, we included outlier volumes defined in the ART movement correction and six movement parameters from the rigid body transformation defined in the realignment procedure as covariates of no interest. At the second level, population-level inferences were based on a random effects model that estimated the second-level statistic at each voxel. Within-group effects were estimated using one-sample t-tests across the single-participant contrast images. For between-group analyses, we used two-sample t-tests comparing the means of the single-subject contrast images from both groups. To control for task difficulty

differences and group differences in the behavioral visual-speech recognition performance, we included the difference between correct responses in the visual-speech task and in the face-identity task as a covariate of no interest into the second-level analysis.

Correlation analyses with behavioral performance

To assess the behavioral relevance of the dorsal–ventral functional connectivity during visual-speech recognition, we performed correlation analyses using SPM12. We entered the visual-speech task score of each participant as a covariate of interest into the second-level analysis.

In the ASD group, we additionally entered a behavioral measure of face recognition abilities as a covariate of no interest into the design. We did this to account for the potential effect of face recognition deficits in ASD (for review, see Weigelt, Koldewyn, & Kanwisher, 2012) on the expected positive correlation between the dorsal–ventral connectivity and visual-speech recognition accuracy. First, we tested which behavioral measure of face recognition abilities (CFPT, CFMT, or a combination of both) predicted visual-speech recognition accuracy best. Second, the visual-speech recognition accuracy score and the face recognition score of choice were orthogonalized using Gram-Schmidt algorithm (Dukes, 2014) to circumvent the problem of multicollinearity due to entering covariates into the design that are correlated (Omidikia & Kompany-Zareh, 2013). Third, dorsal–ventral connectivity and the orthogonalized measures of visual-speech recognition and face recognition were entered into the SPM design. We do not report the r values as an estimation of the effect size of a correlation, because SPM does not provide r values.

Statistical significance thresholds for fMRI results

We conducted a ROI analysis and formulated a-priori hypotheses for four brain regions of the ventral-form pathway (right OFA, left OFA, right FFA, and left FFA) based on the model by Bernstein and Yovel (2015). A hypothesis-driven ROI analysis approach is particularly suitable for studies conducted with relatively small sample sizes to tackle the potential power problem (Cremers, Wager, & Yarkoni, 2017). Effects were considered significant at $p < .05$ corrected for family wise error (FWE) for the ROI. In addition, we applied the Holm–Bonferroni method to correct for multiple comparisons for the four ROIs (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 & Hayasaka, 2003). Effects outside the ROIs were considered significant at $p < .05$ FWE corrected for the whole brain.

Control analyses

We conducted two control analyses. First, we repeated the second-level ROI analysis using the "first-step group" ROIs in the bilateral FFA to check whether the potential effects in these regions can be replicated with a different ROI definition approach including the whole participant sample. Second, we examined if the performance IQ that was significantly different between the ASD and the control group was associated with the dorsal–ventral functional connectivity. For this purpose, we conducted correlation analyses using SPM12. For

each seed region separately, we entered the performance IQ score of each participant as a covariate of interest into the second-level analysis including participants from both groups. We corrected for the four target ROIs using Holm–Bonferroni correction (p < .0125 FWE corrected for the ROI).

3 | RESULTS

3.1 | Behavioral characterization of the study sample

3.1.1 | Visual-speech and face-identity recognition performance and eye tracking during fMRI

The behavioral results of the fMRI visual-speech recognition experiment and the eye tracking results were reported previously (Borowiak et al., 2018). In the context of the present research question, it is important that the ASD group was significantly impaired at visual-speech and face-identity recognition compared to the control group (Figure 2a, Table 2). The eye tracking revealed that the control and the ASD group had similar gaze behavior during the fMRI visual-speech recognition experiment and the fMRI ROI localizer (Table S5).

3.1.2 | Behavioral assessment of facial form perception and facial memory

For the CFPT, a repeated measures ANOVA with the within-subject factor *Orientation* (upright, inverted) and the between-subject factor

Group (control, ASD) revealed a significant main effect of *Orientation* ($F[1,32]$ = 147.708, p = .000, η^2 = .822) indicating that both groups performed significantly better on trials with upright compared to inverted faces. There was a marginally significant main effect of *Group* ($F[1,32]$ = 4.155, p = .050, η^2 = .115), but no significant interaction *Orientation* × *Group* ($F[1,32]$ = 1.958, p = .171, η^2 = .058). Post hoc analyses using Welch's independent-samples *t*-test revealed that compared to the controls, the ASD group performed significantly worse on trials with upright faces ($t[28]$ = 2.493, p = .019, Cohen's d = 0.855), but not on trials with inverted faces ($t[29]$ = 1.264, p = .216, Cohen's d = 0.433; Figure 2b, Table 2).

For the CFMT, we found a significant group difference in facial memory accuracy ($t[30]$ = 2.595, p = .014, Cohen's d = 0.890), because the ASD group had significantly fewer correct responses than the control group (Figure 2c, Table 2).

3.2 | Dorsal-movement and ventral-form regions are functionally connected during visual-speech recognition in typically developing individuals

In the control group, there was functional connectivity between all the seed regions (Figure 3a) of the dorsal-movement pathway and all the target regions in the ventral-form pathway (p < .05 FWE-corrected for the ROI, Figure 3b [green], Table 3). All effects remained significant after Holm–Bonferroni correction for the four ROIs (p < .0125 FWE corrected for the ROI). These results confirmed our hypothesis that dorsal-movement and ventral-form regions are functionally connected during visual-speech recognition in typically developing individuals.
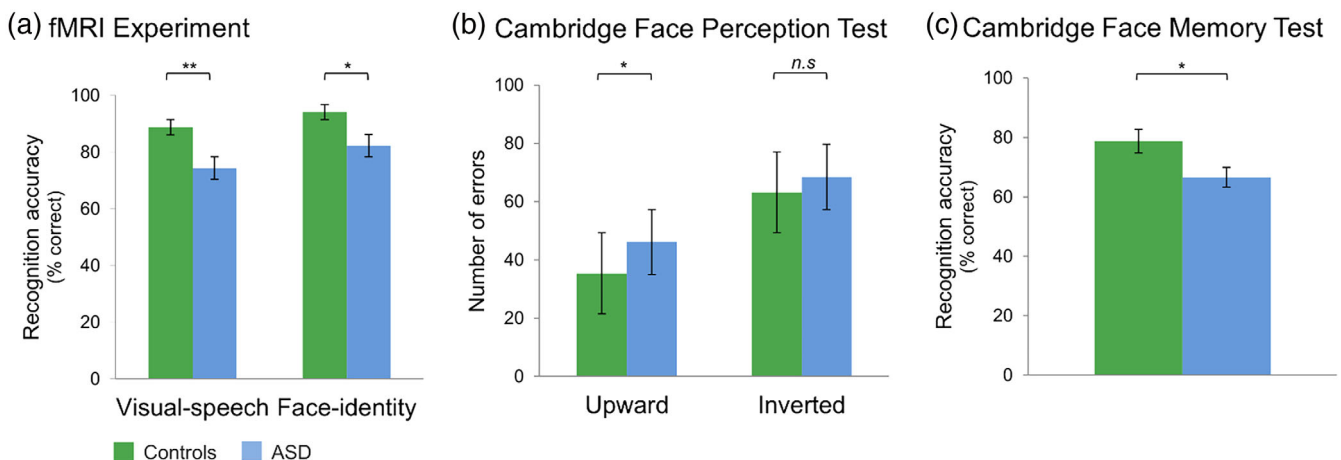

FIGURE 2 Behavioral performance of the ASD and the control group in tests on visual-speech and face recognition. (a) The ASD group performed significantly worse than the control group in the visual-speech task and in the face-identity task. An ANOVA revealed a significant main effect of *Group* and *Task* indicating that the ASD group performed worse on both tasks. Within group comparisons showed that both groups performed significantly worse in the visual-speech task compared to the face-identity task. A score of 50% signified chance performance. (b) The ASD group was significantly worse than the control group in discriminating faces when the faces were presented upright but performed equally well as the control group when the faces were inverted. A score of 93.3 (for upright or inverted trials) signified chance performance. (c) The ASD group also performed significantly worse on face-identity recognition when they were presented with static faces. A score of 33% signified chance performance. Error bars represent ±1SE; ** p < .001; * p < .05, n.s. = not significant [Color figure can be viewed at wileyonlinelibrary.com]

TABLE 2 Summary of average performance scores for visual-speech and face-identity recognition experiments

	Control (n = 17)		ASD (n = 17)		
	M	*SD*	*M*	*SD*	*p*
Visual-speech recognition experiment (recognition accuracy %)					
Visual-speech	88.46	4.48	76.30	11.03	.000*
Face-identity	93.99	6.11	84.63	12.52	.011*
Perception of facial form (CFPT[a]; number of errors)					
Upright faces	33.18	12.33	46.00	17.25	.019*
Inverted faces	62.23	13.19	69.06	17.96	.216
Recognition of face identity (CFMT[b]; recognition accuracy %)					
Total	78.76	12.39	66.50	15.02	.014*
Same images	98.04.	4.79	87.91	15.86	.021*
Different images	79.22	14.98	63.79	16.22	.007*
Different images with noise	63.73	17.73	55.63	16.79	.188

[a]CFPT, Cambridge Face Perception Test (Duchaine et al., 2007).
[b]CFMT, Cambridge Face Memory Test (Duchaine & Nakayama, 2006).
*Significant group differences (*p* < .05); *M* = mean; *SD* = standard deviation.

3.3 | Dorsal–ventral functional connectivity is not correlated with visual-speech recognition performance in typically developing individuals

We conducted a correlation analysis between dorsal–ventral connectivity and recognition accuracy in the visual-speech task. For each seed region, we computed four correlations for which we corrected using Holm–Bonferroni correction (*p* < .0125 FWE corrected for the ROI). We did not find any significant correlations (*p* > .048 uncorrected). These results reject our hypothesis that dorsal–ventral connectivity is positively correlated with visual-speech recognition performance in typically developing individuals.

3.4 | Parts of dorsal–ventral connectivity are reduced in ASD compared to typically developing controls

Similar to the control group, the ASD group had functional connectivity between all the regions in the dorsal-movement (Figure 3a) and the ventral-form pathways (*p* < .05 FWE corrected for the ROI; Figure 3b [blue], Table 3). Most effects remained significant when we applied a Holm–Bonferroni correction for the four ROIs (*p* < .0125 FWE corrected for the ROI), except for right V5/MT—left OFA, right V5/MT—right FFA and right V5/MT—left FFA functional connectivity.

Group comparisons revealed significantly lower functional connectivity in the ASD group compared to the control group between (a) the right V5/MT and the right OFA (*p* < .011 FWE corrected for the ROI), and between (b) the left TVSA and the left FFA (*p* < .012 FWE corrected for the ROI; Figure 3b [purple], Figure S2, Table 3). The group differences remained significant after Holm–Bonferroni correction for four comparisons (*p* < .0125 FWE corrected for the ROI). There was also other functional connectivity that was reduced in the ASD group compared to the

control group (right V5/MT—right FFA, left V5/MT—right FFA), but the group differences did not remain significant after Holm–Bonferroni correction for four comparisons (*p* < .05 FWE corrected for the ROI, Table 3). These results were in partial agreement with both hypotheses that functional connectivity between dorsal-movement and ventral-form regions (a) is partially intact during visual-speech recognition in ASD and that (b) parts of it (right V5/MT—right OFA, left TVSA—left FFA) are reduced compared to typically developing controls.

3.5 | Dorsal–ventral functional connectivity was not significantly correlated with visual-speech recognition performance in ASD

Correlation analysis in the ASD group was conducted in two steps. This was done because of evidence on face recognition deficits in ASD (Weigelt et al., 2012) suggesting that facial form processing might be impaired in ASD. We assume that facial form cues can only be informative for visual-speech recognition, if its processing is intact (Campbell, 1996c; de Gelder & Vroomen, 1998). Therefore, we aimed to account for the potential influence of face recognition deficits on the association between dorsal–ventral functional connectivity and visual-speech recognition performance in ASD. In the first step, we assessed whether visual-speech recognition accuracy was predicted by any face recognition ability measure (CFPT, CFMT, or a combination of both) using linear regression. In the second step, we conducted a correlation analysis between dorsal–ventral functional connectivity and visual-speech recognition accuracy and included the face recognition measure that predicted visual-speech recognition accuracy best as a covariate of no interest.

First, we computed three linear regressions to test whether visual-speech recognition accuracy in the ASD group could be predicted by performance in the (a) CFPT task on upright trials, (b) CFMT task, or (c) by a combination of the two scores. For the
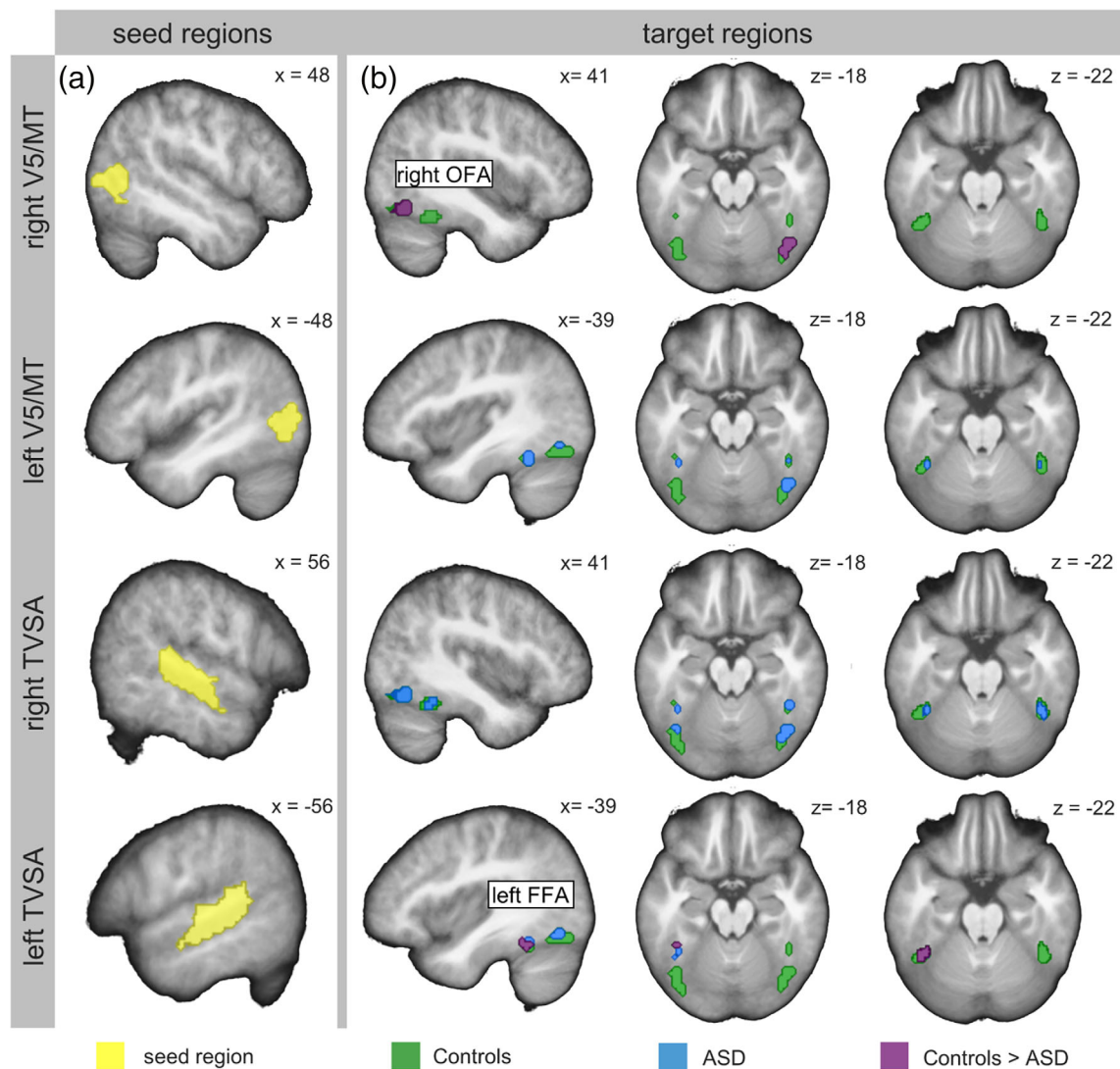
FIGURE 3 Functional connectivity of seed regions in the bilateral TVSA and in the bilateral V5/MT during visual-speech recognition. (a) Seed regions were extracted within a sphere of 4 mm around each subject's individual peak coordinate for the contrast "visual-speech task > face-identity task" for each individual participant. Seed regions were located within the anatomical probabilistic map of the V5/MT and the pSTS/G for TVSA (V5/MT: Jülich histological [cyto- and myelo-architectonic] atlas [Eickhoff et al., 2007]; the pSTS/STG: Harvard–Oxford cortical structural atlas [Desikan et al., 2006]). (b) The movement-sensitive bilateral TVSA and the bilateral V5/MT were functionally connected to the form-sensitive regions in the ventral pathway (OFA, FFA) in the control group (green; $p \leq .023$ FWE corrected, Holm–Bonferroni corrected), and in the ASD group (blue; $p \leq .040$ FWE corrected, Holm–Bonferroni corrected). The control group showed higher functional connectivity than the ASD group between the right V5/MT and the right OFA (purple; $p = .011$ FWE corrected, Holm–Bonferroni corrected) and the left TVSA and the left FFA (purple; $p = .012$ FWE corrected, Holm–Bonferroni corrected). For display purposes within-group effects are presented at the threshold of $p = .005$ 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. TVSA, temporal visual speech area; V5/MT, visual area 5/middle temporal area; FFA, fusiform face area; OFA, occipital face area; x, z, MNI-coordinates [Color figure can be viewed at wileyonlinelibrary.com]

CFPT score, a marginally significant regression equation was found with an R^2 of .232 and an adjusted R^2 of .181 ($F[1,15] = 4.540$, $p = .050$). For the CFMT score, a significant regression equation was found with an R^2 of .345 and an adjusted R^2 of .302 ($F[1,15] = 7.917$, $p = .013$). A regression model including both the CFPT score and the CFMT score resulted in R^2 of .364 and an adjusted R^2 of .273 ($F[1,15] = 4.008$, $p = .042$). The regression analysis demonstrated that the CFMT score alone was the best predictor for the visual-speech

recognition performance in the ASD group (the highest adjusted R^2). The visual-speech recognition accuracy score and the CFMT score were orthogonalized using Gram-Schmidt algorithm to circumvent the problem of multicollinearity.

Second, we conducted correlation analyses between the dorsal–ventral connectivity and recognition accuracy in the visual-speech task and included the orthogonalized CFMT task score as a covariate of no interest into the correlation analyses. For each seed region, we

computed four correlations for which we corrected using Holm–Bonferroni correction ($p < .0125$ FWE corrected for the ROI). We found a positive correlation between the visual-speech task performance and functional connectivity between the left V5/MT—right FFA ($p = .030$ FWE corrected for the ROI) and between the left V5/MT and the left OFA ($p = .045$ FWE corrected for the ROI). However, none of the correlations remained significant after Holm-Bonferroni correction for four calculations. For the remaining functional connectivity, we did not find any significant correlations ($p > .014$ uncorrected).

TABLE 3 MNI coordinates of ventral-form regions that showed functional connectivity to dorsal-movement regions during the visual-speech task compared to the face-identity task

Seed region: Right V5/MT

Region		Control x	y	z	Z	p	ASD x	y	z	Z	p
OFA	r	45	−67	−19	5.13	.000	42	−67	−13	3.51	.004
		39	−70	−13	5.10	.000					
		39	−64	−13	4.95	.000					
		39	−79	−16	4.30	.000					
	l	−39	−70	−19	4.16	.000	−36	−79	−13	2.87	.026
		−39	−73	−13	3.96	.001					
		−36	−82	−13	3.31	.004					
		−33	−85	−16	2.92	.023					
FFA	r	42	−49	−25	4.14	.000	45	−58	−25	2.47	.046
	l	−45	−55	−22	3.94	.001	−39	−52	−19	2.81	.023
		−39	−49	−25	3.93	.001					
		Control > ASD					**ASD > Control**				
OFA	r	39	−64	−16	3.11	.011	−				
FFA	r	42	−46	−25	2.65	.028	−				

Seed region: Left V5/MT

Region		Control x	y	z	Z	p	ASD x	y	z	Z	p
OFA	r	39	−67	−16	4.14	.000	42	−67	−13	3.76	.001
							36	−79	−13	3.37	.005
	l	−42	−73	−16	3.52	.004	−36	−76	−10	3.94	.001
		−36	−79	−13	3.09	.008					
FFA	r	42	−46	−19	4.40	.000	42	−52	−22	3.05	.010
	l	−42	−46	−19	3.91	.001	−39	−52	−19	3.43	.004
		−42	−55	−22	3.24	.006					
		−39	−49	−22	3.21	.007					
		Control > ASD					**ASD > Control**				
FFA	r	42	−46	−25	2.62	.028	−				

Seed region: Right TVSA

Region		Control x	y	z	Z	p	ASD x	y	z	Z	p
OFA	r	39	−67	−16	4.81	.000	39	−73	−13	3.64	.003
		39	−73	−13	4.69	.000	36	−79	−13	3.46	.005
		36	−79	−13	4.38	.000					
	l	−39	−70	−19	4.30	.000	−36	−79	−10	3.76	.002
		−33	−85	−16	3.56	.002	−33	−85	−16	3.20	.011
FFA	r	42	−46	−19	4.42	.000	42	−46	−19	3.91	.001
		45	−55	−22	3.60	.002					

(Continues)

TABLE 3 (Continued)

Seed region: Right TVSA										
		Control					**ASD**			
	l	−45	−55	−25	4.12	.000	−39	−52	−22	3.02	.014
		−42	−46	−19	3.80	.001				
		−39	−49	−25	3.73	.001				
		Control > ASD					**ASD > Control**			
		–					–			

Seed region: Left TVSA										
		Control					**ASD**			
OFA	r	39	−64	−13	5.67	.000	42	−64	−13	3.31	.007
	l	−36	−76	−19	4.06	.001	−36	−76	−10	3.40	.006
		−39	−67	−16	4.05	.001				
		−42	−73	−16	4.05	.001				
		−36	−82	−13	3.76	.002				
FFA	r	42	−46	−22	5.85	.000	42	−49	−19	2.52	.040
	l	−39	−46	−22	5.12	.000	−39	−52	−19	2.98	.014
		Control > ASD					**ASD > Control**			
FFA	l	−39	−46	−22	2.98	.012	–			

Note: Coordinates represent local connectivity maxima in MNI space (in mm) for the whole brain. Clusters reported in normal font reached significance at $p < .05$ FWE (peak-level) corrected for the respective ROI, and remained significant after Holm-Bonferroni correction for the four ROIs. Coordinates written in italics represent clusters that reached significance at $p < .05$ FWE corrected (peak-level) corrected for the respective ROI, but did not remain significant after Holm-Bonferroni correction for the four ROIs. 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 myelo-architectonic] atlas; [Eickhoff et al., 2007]) implemented in FSL (Smith et al., 2004, http://www.fmrib.ox.ac.uk/fsl/fslview). TVSA, temporal visual speech area; V5/MT, visual area 5/middle temporal area; FFA, fusiform face area; OFA, occipital face area.

3.6 | Control analyses

First, the second-level ROI analysis using "first-step group" ROIs in the bilateral FFA confirmed the results of the main second-level ROI analysis using "second-step group" ROIs (see Supporting Information Results, Table S6). Second, there was no significant correlation between the performance IQ scores and the dorsal-ventral functional connectivity to any of the four target ROIs across both groups (all values of $p \geq .067$ FWE corrected for the ROI). Therefore, the control analyses demonstrated that (a) the reported effects can be replicated using a different ROI definition approach and that (b) the group differences in parts of the dorsal–ventral functional connectivity are likely not influenced by group differences in performance IQ.

3.7 | Summary of results

A schematic overview of the dorsal–ventral functional connectivity is displayed in Figure 4a for the control group and in Figure 4b for the ASD group. Figure 4b additionally displays the differences found between the ASD group and the control group in brain responses and functional connectivity along the dorsal-movement

pathway (Borowiak et al., 2018) and in functional connectivity between the dorsal-movement and the ventral-form pathway (present study).

4 | DISCUSSION

Our study revealed two main findings. First, visual cortex regions in the dorsal-movement and in the ventral-form pathway were functionally connected to each other during the recognition of visual speech, in contrast to face identity. This was the case in both typically developing adults and adults with ASD. Second, parts of the dorsal–ventral functional connectivity were reduced in the ASD group compared to the control group (i.e., right V5/MT–right OFA, left TVSA—left FFA). The contribution of the results is twofold. First, they broaden our knowledge about dynamic face perception by showing that functional connectivity between dorsal-movement and ventral-form regions exists not only for facial emotion (Foley et al., 2012; Furl et al., 2014), but also for visual-speech processing. This corroborates the revised theoretical model for dynamic face perception by Bernstein and Yovel (2015) by showing that perception of dynamic facial information involves a brain network consisting of regions not only in the dorsal-
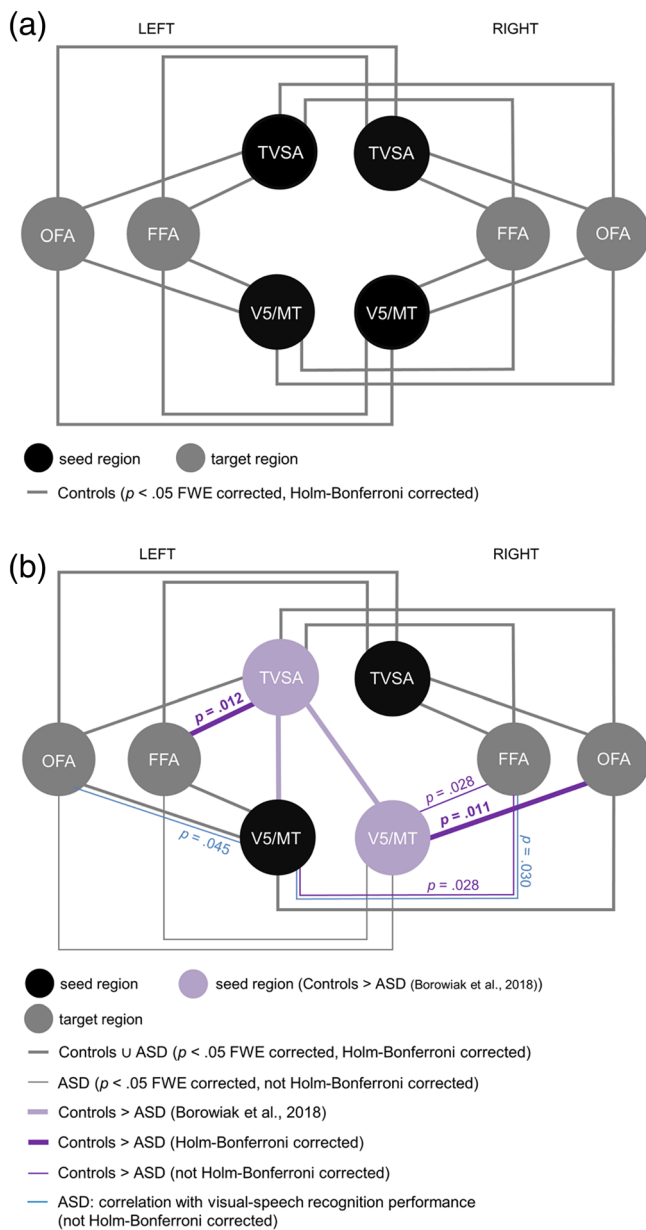
FIGURE 4 Overview of functional connectivity patterns between the dorsal-movement bilateral TVSA and V5/MT (seed regions; black and purple) and ventral-form regions in the bilateral FFA and OFA (target regions; gray). (a) Functional connectivity in the typically developing control group. (b) Functional connectivity in the ASD group, in comparison to the typically developing control group. Seed regions that are marked in black have been shown to have comparable local brain responses to visual-speech recognition between ASD and typically developing individuals. Seed regions that are marked in purple have been shown to have reduced local brain responses to visual-speech recognition in ASD compared to typically developing individuals (Borowiak et al., 2018). For information purposes, (b) displays also the functional connectivity in the ASD group and the group differences between the control and the ASD group that did not survive Holm–Bonferroni correction [Color figure can be viewed at wileyonlinelibrary.com]

movement pathway, but also in the ventral-form pathway. Second, dysfunction in parts of the dorsal–ventral network might represent a mechanism that may contribute to difficulties in processing dynamic

facial information in ASD (e.g., O'Brien, Spencer, Girges, Johnston, & Hill, 2014; Sato et al., 2013; Schelinski et al., 2014).

The finding that there is dorsal–ventral functional connectivity for visual-speech, similar to facial emotion perception, sheds new light on the perceptual mechanisms behind visual-speech recognition. Facial emotion and visual-speech processing are primarily reliant on movement information and can be recognized from point-light displays without any form information (for review, see Blake & Shiffrar, 2007). This suggests a main involvement of the dorsal-movement pathway. However, there are clinical studies, which report that patients with lesions in the ventral temporal lobe show deficits in processing facial emotion and visual speech despite intact dorsal-movement regions (Barton, 2008; Campbell, 1996c; de Gelder & Vroomen, 1998; Humphreys, Avidan, & Behrmann, 2007). This indicates that intact processing of movement alone may not be sufficient to accomplish recognition of facial emotion and visual speech. Perceiving facial expressions relies on a mixture of holistic and feature-based componential processes (Calder, Young, Keane, & Dean, 2000; Palermo et al., 2011; Tanaka, Kaiser, Butler, & Le Grand, 2012), that might require functional interactions of dorsal-movement and ventral-form regions (Foley et al., 2012; Furl et al., 2014). Visual-speech recognition involves feature-based processing of the dynamic face, because seeing only the mouth area including lips, tongue, teeth, and mandible is sufficient for recognition (Marassa & Lansing, 1995; Stone, 1957; Thomas & Jordan, 2004). Form information from the relevant face features is likely processed in regions of the ventral-form pathway and integrated with information processed in the dorsal-movement pathway via the dorsal–ventral functional connectivity. Evidence for the recruitment of dorsal–ventral functional connectivity during visual-speech recognition provides a mechanism for how form information might contribute to visual-speech recognition.

The behavioral relevance of the dorsal–ventral functional connectivity for visual-speech recognition remains an open question. Clinical case studies reported that patients with lesions in the ventral temporo-occipital and/or with visual form agnosia showed deficits in visual-speech recognition (e.g., Campbell, 1992; Campbell, 1996b; Campbell et al., 1986; Campbell et al., 1990). Moreover, fusiform gyrus responses to visual speech versus a baseline static face condition correlated with visual-speech recognition performance (Capek et al., 2008). These findings suggested that ventral-form regions for face perception might be involved in recognizing speech from faces, at least to some degree (Campbell, 2011). We propose that dorsal–ventral connectivity is responsive to visual-speech recognition and that this network may function as a complementary system for visual-speech processing in dorsal-movement regions. Such a mechanism could be recruited to promote recognition, when visual-speech information is particularly complex or difficult to recognize by providing additional form information from the face. In our study, we did not find any correlation with visual-speech recognition behavior in the typically developing controls. A potential reason for this might be that the visual-speech recognition task was sufficiently easy to accomplish by relying on the primary dorsal-movement pathway alone. Such an interpretation is consistent with the relatively high recognition

accuracy (88%) observed in the visual-speech task in the control group. Investigating the behavioral relevance of the dorsal–ventral functional connectivity might require visual-speech recognition tasks with a higher difficulty level and a higher performance variability to detect an association between dorsal–ventral functional connectivity and behavioral performance.

In ASD, multiple studies have shown that brain responses and functional connectivity are reduced in the dorsal-movement pathway (i.e., V5/MT and the pSTS/STG [TVSA]) during general and human movement perception (e.g., Alaerts et al., 2013; Alaerts, Swinnen, & Wenderoth, 2017; Freitag et al., 2008; Herrington et al., 2007; Sato et al., 2012). The present ASD study sample had reduced brain responses and functional connectivity between the right V5/MT and the left TVSA during visual-speech recognition, compared to typically developing controls (Borowiak et al., 2018). In this study, we showed that the same regions have also reduced functional connectivity to ventral-form regions involved in facial form processing. Interestingly, other dorsal-movement regions, which had intact responses to visual speech (left V5/MT, right TVSA), showed functional connectivity to the ventral-form pathway that was comparable to the typically developing controls. The present finding of the reduced functional connectivity is likely not due to the decreased responses in the dorsal-movement pathway, because PPI analysis allows assessing task-specific changes in the relationship between responses in different brain regions independent of their amplitude (O'Reilly et al., 2012). In addition, both groups were pairwise-matched on the full-scale IQ and the performance IQ was not significantly associated with the dorsal–ventral functional connectivity. This indicates that the reduced dorsal–ventral functional connectivity in ASD could not be explained by differences in the performance IQ between the groups (Lawrence et al., 2008; Lawrence et al., 2015).

It is difficult to adjudicate where this reduced dorsal–ventral functional connectivity stems from. Functional and structural evidence suggests that deficits in the dorsal-movement pathway for visual-movement perception might be central to the difficulties in dynamic face processing observed in ASD. First, the functional abnormalities in the sensory V5/MT and TVSA during the perception of general and human movement stand in contrast to intact processing in higher-order non-perceptual regions (Borowiak et al., 2018; Robertson et al., 2014). The reduced right V5/MT responses were associated with lower performance in the visual-speech recognition task in ASD indicating their behavioral relevance for the deficit (Borowiak et al., 2018). Second, in the neurotypical population, the V5/MT has a central position in the dorsal–ventral network (Furl, 2015). The V5/MT is structurally well connected to both the dorsal–movement pSTS/STG (Bernstein, Erez, Blank, & Yovel, 2018; Ethofer et al., 2011), and the ventral-form OFA and the FFA (Bernstein et al., 2018; Ethofer et al., 2011; Kim et al., 2006). In contrast, previous studies showed low structural connectivity between the voice-sensitive portion of the pSTS/STG and the FFA (Blank, Anwander, & von Kriegstein, 2011), and between the face-sensitive pSTS/STG and the FFA and the OFA (Gschwind et al., 2012; Pyles et al., 2013). To our knowledge, structural connectivity of the TVSA has not been specifically investigated

yet. Based on these findings, we speculate that alterations of the right V5/MT and its functional connectivity to the left TVSA within the dorsal-movement pathway (Borowiak et al., 2018) might have a two-fold implication for the dorsal–ventral network. First, dysfunction of the right V5/MT might likely lead to reductions in functional connectivity to regions that it is directly connected to, such as the OFA (as observed in the present study). Second, it could indirectly influence the functional connectivity between the left TVSA and ventral-form regions due to its central position as a connecting region. In this context, impairments of the dorsal visual-movement regions in ASD would affect not only the processing of movement information via the V5/MT—TVSA (i.e., pSTS/STG) functional connectivity, but also the extraction of form information from the movement via the functional connectivity between the V5/MT and the OFA and the FFA (O'Toole et al., 2002). This suggests that functional alterations of the dorsal-movement pathway could also contribute to difficulties with identity recognition from the dynamic face, a deficit that is consistently observed in in ASD (O'Brien et al., 2014), and also present in this ASD study sample. However, such an assumption is still speculative, as functional connectivity analysis does not reveal the directionality of information flow between brain regions (Friston et al., 1997).

One prerequisite for a complementary role of the dorsal–ventral functional connectivity for visual-speech recognition is that facial form information provided by the network connectivity is informative. However, face recognition difficulties known in ASD suggest that processing of facial form might be deficient. In our study, behavioral assessment of facial form perception and facial memory demonstrated significantly reduced performance in ASD compared to typically developing adults. The ability to memorize faces significantly predicted visual-speech recognition accuracy in ASD. Therefore, we regressed out its effects in the correlation analyses of the dorsal–ventral functional connectivity and visual-speech recognition performance. This approach revealed behavioral relevance of the left V5/MT functional connectivity to the ventral-form regions for visual-speech recognition in ASD suggesting that the dorsal–ventral functional connectivity might be recruited when visual-speech recognition becomes more challenging. However, the positive correlation did not survive the correction for multiple comparisons and should be therefore interpreted with caution. We speculate that the behavioral tasks for face recognition abilities used in our study might not have been specifically targeted toward examining the facial form mechanisms relevant for visual-speech recognition. For example, tests on perception of facial parts might be more related to visual speech (Lansing & McConkie, 2003), while perception of whole faces is likely more relevant for identity recognition (Farah et al., 1998). In addition, behavioral tasks that would include dynamic face stimuli might be more natural and ecologically valid as visual speech is dynamic by nature. Implementing such assessments might also allow an examination of the process of extracting facial form information from facial movement, a task which might also recruit dorsal–ventral functional connectivity (O'Toole et al., 2002).

So far, only two other studies investigated network connectivity during visual-speech recognition in typically developing individuals. These studies demonstrated functional connectivity between dorsal-

movement regions and other visual-speech related regions (Borowiak et al., 2018; Chu et al., 2013). However, neither study aimed to specifically investigate network connectivity between the dorsal-movement and the ventral-form regions. In the present study, we functionally defined the face-sensitive FFA, OFA, and the face-movement sensitive portion of the pSTS/STG (i.e., TVSA) using an independent functional localizer and functional probabilistic maps for face-sensitive regions. This is crucial because the fusiform gyrus is functionally heterogeneous and comprises regions that can be classified according to their preferential responses to faces (e.g., Kanwisher et al., 1997), house and places (e.g., Aguirre, Zarahn, & D'Esposito, 1998; Epstein & Kanwisher, 1998), and human movement (e.g., Grossman et al., 2000; Puce et al., 1998). Similarly, the posterior STS/STG is responsive to visual (e.g., Beauchamp, Lee, Haxby, & Martin, 2002; Grossman et al., 2000), auditory (e.g., Fecteau, Armony, Joanette, & Belin, 2004; von Kriegstein & Giraud, 2004), and audio–visual stimuli (e.g., Beauchamp, Argall, Bodurka, Duyn, & Martin, 2004; Wright, Pelphrey, Allison, McKeown, & McCarthy, 2003). Chu et al. (2013) reported functional connectivity between regions labeled as the lateral posterior fusiform gyrus and the posterior STG by comparing visual-speech perception to a baseline condition. However, on a closer look, the reported MNI coordinates (x = 50, y = −50, z = 0; x = −49, y = −59, z = 0) did not correspond to the typical functional location of the FFA or the OFA (Blank et al., 2014; Sabatinelli et al., 2011). In addition, brain responses to visual speech along the fusiform gyrus have been reported in studies using less specific contrasts between visual speech and static face images or gurning (Calvert & Campbell, 2003; Campbell et al., 2001; Capek et al., 2008). Here, we were able to more specifically target mechanisms underlying visual-speech processing in contrast to processing of other face information by comparing the visual-speech task to the face-identity task.

In conclusion, the present study revealed that functional connectivity between dorsal-movement and ventral-form regions exists also during visual-speech recognition in the typically developing population and in ASD, but parts of it are reduced in ASD. Impairments of the dorsal–ventral functional connectivity in ASD can be observed for visual-speech perception, but might also exist during processing of other socially relevant information in the dynamic face and thus contribute to face-to-face communication difficulties typical for ASD.

ACKNOWLEDGMENTS

CONFLICT OF INTEREST

The authors declare no competing financial interest.

AUTHOR CONTRIBUTIONS

K.B. and K.v.K. designed research; K.B. performed research; K.B. and C.M. performed data analysis; K.B. and K.v.K. wrote the article; C.M. contributed to article writing.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

ORCID

Kamila Borowiak https://orcid.org/0000-0002-0216-5613
Corrina Maguinness https://orcid.org/0000-0002-6200-4109
Katharina von Kriegstein https://orcid.org/0000-0001-7989-5860

REFERENCES

Aguirre, G. K., Zarahn, E., & D'Esposito, M. (1998). An area within human ventral cortex sensitive to "building" stimuli: Evidence and implications. *Neuron*, 21, 373–383.

Alaerts, K., Swinnen, S. P., & Wenderoth, N. (2017). Neural processing of biological motion in autism: An investigation of brain activity and effective connectivity. *Scientific Reports*, 7, 5612.

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. *Social Cognitive and Affective Neuroscience*, 9, 1589–1600.

Allison, T., Puce, A., & McCarthy, G. (2000). Social perception from visual cues: Role of the STS region. *Trends in Cognitive Sciences*, 4, 267–278.

American Psychiatric Association [APA] (2013). *Diagnostic and statistical manual of mental disorders (DSM-5)* (5th edn.). Washington, DC: American Psychiatric Association.

Andrews, T. J., & Ewbank, M. P. (2004). Distinct representations for facial identity and changeable aspects of faces in the human temporal lobe. *NeuroImage*, 23, 905–913.

Anzellotti, S., & Caramazza, A. (2017). Multimodal representations of person identity individuated with fMRI. *Cortex*, 89, 85–97.

Arnold, P., & Hill, F. (2001). Bisensory augmentation: A speechreading advantage when speech is clearly audible and intact. *British Journal of Psychology*, 92, 339–355.

Aschenberner, B., & Weiss, C. (2005). *Phoneme-viseme mapping for German video-realistic audio-visual-speech-synthesis*. IKP-working paper NF11.

Ashwood, K. L., Gillan, N., Horder, J., Hayward, H., Woodhouse, E., McEwen, F. S., … Cadman, T. (2016). Predicting the diagnosis of autism in adults using the autism-Spectrum quotient (AQ) questionnaire. *Psychological Medicine*, 46, 2595–2604.

Avidan, G., & Behrmann, M. (2014). Impairment of the face processing network in congenital prosopagnosia. *Frontiers in Bioscience (Elite Edition)*, 6, 236–257.

Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J., & Clubley, E. (2001). The autism-Spectrum quotient (AQ): Evidence from Asperger syndrome/high-functioning autism, males and females, scientists and mathematicians. *Journal of Autism and Developmental Disorders*, 31, 5–17.

Barton, J. J. (2008). Prosopagnosia associated with a left occipitotemporal lesion. *Neuropsychologia*, 46, 2214–2224.

Beauchamp, M. S., Argall, B. D., Bodurka, J., Duyn, J. H., & Martin, A. (2004). Unraveling multisensory integration: Patchy organization within human STS multisensory cortex. *Nature Neuroscience*, 7, 1190.

Beauchamp, M. S., Lee, K. E., Haxby, J. V., & Martin, A. (2002). Parallel visual motion processing streams for manipulable objects and human movements. *Neuron*, 34, 149–159.

Beckers, G., & Homberg, V. (1992). Cerebral visual motion blindness: Transitory akinetopsia induced by transcranial magnetic stimulation of human area V5. *Proceedings of the Royal Society of London B: Biological Sciences*, 249, 173–178.

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. *Human Brain Mapping*, 32, 1660–1676.

Bernstein, M., Erez, Y., Blank, I., & Yovel, G. (2018). An integrated neural framework for dynamic and static face processing. *Scientific Reports*, *8*, 7036.

Bernstein, M., & Yovel, G. (2015). Two neural pathways of face processing: A critical evaluation of current models. *Neuroscience and Biobehavioral Reviews*, *55*, 536–546.

Blake, R., & Shiffrar, M. (2007). Perception of human motion. *Annual Review of Psychology*, *58*, 47–73.

Blank, H., Anwander, A., & von Kriegstein, K. (2011). Direct structural connections between voice-and face-recognition areas. *Journal of Neuroscience*, *31*, 12906–12915.

Blank, H., Wieland, N., & von Kriegstein, K. (2014). Person recognition and the brain: Merging evidence from patients and healthy individuals. *Neuroscience and Biobehavioral Reviews*, *47*, 717–734.

Bölte, S., Rühl, D., Schmötzer, G., & Poustka, F. (2003). *Diagnostisches Interview für Autismus—Revidiert (ADI-R)*. Bern: Verlag Hans Huber.

Borowiak, K., Schelinski, S., & von Kriegstein, K. (2018). Recognizing visual speech: Reduced responses in visual-movement regions, but not other speech regions in autism. *NeuroImage: Clinical*, *20*, 1078–1091.

Brickenkamp, R. (2002). *Test d2—Aufmerksamkeits-Belastung-Test (d2)*. Göttingen: Hogrefe.

Bruce, V., & Young, A. (1986). Understanding face recognition. *British Journal of Psychology*, *77*, 305–327.

Calder, A. J., Young, A. W., Keane, J., & Dean, M. (2000). Configural information in facial expression perception. *Journal of Experimental Psychology: Human Perception and Performance*, *26*, 527.

Calvert, G. A., & Campbell, R. (2003). Reading speech from still and moving faces: The neural substrates of visible speech. *Journal of Cognitive Neuroscience*, *15*, 57–70.

Campbell, R. (1992). The neuropsychology of lipreading. *Philosophical Transactions of the Royal Society B*, *335*, 39–45.

Campbell, R. (1996a). Dissociating face processing skills: Decisions about lip read speech, expression, and identity. *The Quarterly Journal of Experimental Psychology: Section A*, *49*, 295–314.

Campbell, R. (1996b). Seeing brains reading speech: A review and speculations. In *Speechreading by humans and machines* (pp. 115–133). Berlin, Heidelberg: Springer.

Campbell, R. (1996c). Seeing speech in space and time. *Proc. 4th Internat. Conf. Spoken Language Processing*. Philadelphia, PA, October 1996.

Campbell, R. (2011). Speechreading and the Bruce–Young model of face recognition: Early findings and recent developments. *British Journal of Psychology*, *102*, 704–710.

Campbell, R., Garwood, J., Franklin, S., Howard, D., Landis, T., & Regard, M. (1990). Neuropsychological studies of auditory-visual fusion illusions. Four case studies and their implications. *Neuropsychologia*, *28*, 787–802.

Campbell, R., Landis, T., & Regard, M. (1986). Face recognition and lipreading: A neurological dissociation. *Brain*, *109*, 509–521.

Campbell, R., MacSweeney, M., Surguladze, S., Calvert, G., McGuire, P., Suckling, 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). *Cognitive Brain Research*, *12*, 233–243.

Capek, C. M., MacSweeney, M., Woll, B., Waters, D., McGuire, P. K., David, A. S., ... Campbell, R. (2008). Cortical circuits for silent speechreading in deaf and hearing people. *Neuropsychologia*, *46*, 1233–1241.

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.

Cremers, H. R., Wager, T. D., & Yarkoni, T. (2017). The relation between statistical power and inference in fMRI. *PLoS One*, *12*, e0184923.

Dalton, K. M., Nacewicz, B. M., Johnstone, T., Schaefer, H. S., Gernsbacher, M. A., Goldsmith, H. H., ... Davidson, R. J. (2005). Gaze fixation and the neural circuitry of face processing in autism. *Nature Neuroscience*, *8*, 519.

Davis, J. M., McKone, E., Dennett, H., O'Connor, K. B., O'Kearney, R., & Palermo, R. (2011). Individual differences in the ability to recognise facial identity are associated with social anxiety. *PLoS One*, *6*, e28800.

de Gelder, B., & Vroomen, J. (1998). Impairment of speech-reading in prosopagnosia. *Speech Communication*, *26*, 89–96.

Desikan, R. S., Ségonne, F., Fischl, B., Quinn, B. T., Dickerson, B. C., Blacker, D., ... Albert, M. S. (2006). An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *NeuroImage*, *31*, 968–980.

Dobs, K., Schultz, J., Bülthoff, I., & Gardner, J. L. (2018). Task-dependent enhancement of facial expression and identity representations in human cortex. *NeuroImage*, *172*, 689–702.

Duchaine, B., Germine, L., & Nakayama, K. (2007). Family resemblance: Ten family members with prosopagnosia and within-class object agnosia. *Cognitive Neuropsychology*, *24*, 419–430.

Duchaine, B., & Nakayama, K. (2006). The Cambridge Face Memory Test: Results for neurologically intact individuals and an investigation of its validity using inverted face stimuli and prosopagnosic participants. *Neuropsychologia*, *44*, 576–585.

Dukes, K. A. (2014). Gram–Schmidt process. *Wiley StatsRef: Statistics Reference Online*.

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.

Engell, A. D., & McCarthy, G. (2013). Probabilistic atlases for face and biological motion perception: An analysis of their reliability and overlap. *NeuroImage*, *74*, 140–151.

Epstein, R., & Kanwisher, N. (1998). A cortical representation of the local visual environment. *Nature*, *392*, 598–601.

Ethofer, T., Gschwind, M., & Vuilleumier, P. (2011). Processing social aspects of human gaze: A combined fMRI-DTI study. *NeuroImage*, *55*, 411–419.

Fairhall, S. L., & Ishai, A. (2006). Effective connectivity within the distributed cortical network for face perception. *Cerebral Cortex*, *17*, 2400–2406.

Farah, M. J., Wilson, K. D., Drain, M., & Tanaka, J. N. (1998). What is "special" about face perception? *Psychological Review*, *105*, 482.

Fecteau, S., Armony, J. L., Joanette, Y., & Belin, P. (2004). Is voice processing species-specific in human auditory cortex? An fMRI study. *NeuroImage*, *23*, 840–848.

Foley, E., Rippon, G., Thai, N. J., Longe, O., & Senior, C. (2012). Dynamic facial expressions evoke distinct activation in the face perception network: A connectivity analysis study. *Journal of Cognitive Neuroscience*, *24*, 507–520.

Fox, C. J., Hanif, H. M., Iaria, G., Duchaine, B. C., & Barton, J. J. (2011). Perceptual and anatomic patterns of selective deficits in facial identity and expression processing. *Neuropsychologia*, *49*, 3188–3200.

Foxe, J. J., Molholm, S., Del Bene, V. A., Frey, H. P., Russo, N. N., Blanco, 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. *Cerebral Cortex*, *25*, 298–312.

Freitag, C. M., Konrad, C., Häberlen, M., Kleser, C., von Gontard, A., Reith, W., ... Krick, C. (2008). Perception of biological motion in autism spectrum disorders. *Neuropsychologia*, *46*, 1480–1494.

Friston, K. J., Ashburner, J. T., Kiebel, S. J., Nichols, T. E., & Penny, W. (Eds.) (2007). *Statistical parametric mapping: The analysis of functional brain images*. London, UK: Academic Press.

Friston, K. J. (1994). Functional and effective connectivity in neuroimaging: A synthesis. *Human Brain Mapping*, *2*, 56–78.

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.

Furl, N. (2015). Structural and effective connectivity reveals potential network-based influences on category-sensitive visual areas. *Frontiers in Human Neuroscience*, *9*, 253.

Furl, N., Henson, R. N., Friston, K. J., & Calder, A. J. (2014). Network interactions explain sensitivity to dynamic faces in the superior temporal sulcus. *Cerebral Cortex*, *25*, 2876–2882.

Ganel, T., Valyear, K. F., Goshen-Gottstein, Y., & Goodale, M. A. (2005). The involvement of the "fusiform face area" in processing facial expression. *Neuropsychologia*, *43*, 1645–1654.

Gauthier, I., Skudlarski, P., Gore, J. C., & Anderson, A. W. (2000). Expertise for cars and birds recruits brain areas involved in face recognition. *Nature Neuroscience*, *3*, 191–197.

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.

Grill-Spector, K., Knouf, N., & Kanwisher, N. (2004). The fusiform face area subserves face perception, not generic within-category identification. *Nature Neuroscience*, *7*, 555–562.

Grossman, E., Donnelly, M., Price, R., Pickens, D., Morgan, V., Neighbor, G., & Blake, R. (2000). Brain areas involved in perception of biological motion. *Journal of Cognitive Neuroscience*, *12*, 711–720.

Grossman, E. D., Battelli, L., & Pascual-Leone, A. (2005). Repetitive TMS over posterior STS disrupts perception of biological motion. *Vision Research*, *45*, 2847–2853.

Gschwind, M., Pourtois, G., Schwartz, S., Van De Ville, D., & Vuilleumier, P. (2012). White-matter connectivity between face-responsive regions in the human brain. *Cerebral Cortex*, *22*, 1564–1576.

Harris, A., & Aguirre, G. K. (2010). Neural tuning for face wholes and parts in human fusiform gyrus revealed by FMRI adaptation. *Journal of Neurophysiology*, *104*, 336–345.

Haxby, J. V., Hoffman, E. A., & Gobbini, M. I. (2000). The distributed human neural system for face perception. *Trends in Cognitive Sciences*, *4*, 223–233.

Herrington, J. D., Baron-Cohen, S., Wheelwright, S. J., Singh, K. D., Bullmore, E. T., Brammer, M., & Williams, S. C. (2007). The role of MT +/V5 during biological motion perception in Asperger syndrome: An fMRI study. *Research in Autism Spectrum Disorders*, *1*, 14–27.

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, *6*, 65–70.

Humphreys, K., Avidan, G., & Behrmann, M. (2007). A detailed investigation of facial expression processing in congenital prosopagnosia as compared to acquired prosopagnosia. *Experimental Brain Research*, *176*, 356–373.

Irwin, J. R., & Brancazio, L. (2014). Seeing to hear? Patterns of gaze to speaking faces in children with autism spectrum disorders. *Frontiers in Psychology*, *5*, 397.

Irwin, J. R., Tornatore, L. A., Brancazio, L., & Whalen, D. H. (2011). Can children with autism spectrum disorders "hear" a speaking face? *Child Development*, *82*, 1397–1403.

Jezzard, P., & Balaban, R. S. (1995). Correction for geometric distortion in echo planar images from B0 field variations. *Magnetic Resonance in Medicine*, *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. *Social Cognitive and Affective Neuroscience*, *12*, 319–328.

Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: A module in human extrastriate cortex specialized for face perception. *Journal of Neuroscience*, *17*, 4302–4311.

Kilts, C. D., Egan, G., Gideon, D. A., Ely, T. D., & Hoffman, J. M. (2003). Dissociable neural pathways are involved in the recognition of emotion in static and dynamic facial expressions. *NeuroImage*, *18*, 156–168.

Kim, M., Ducros, M., Carlson, T., Ronen, I., He, S., Ugurbil, K., & Kim, D. S. (2006). Anatomical correlates of the functional organization in the human occipitotemporal cortex. *Magnetic Resonance Imaging*, *24*, 583–590.

Kliemann, D., Richardson, H., Anzellotti, S., Ayyash, D., Haskins, A. J., Gabrieli, J. D., & Saxe, R. R. (2018). Cortical responses to dynamic emotional facial expressions generalize across stimuli, and are sensitive to task-relevance, in adults with and without autism. *Cortex*, *103*, 24–43.

Knappmeyer, B., Thornton, I. M., & Bülthoff, H. H. (2003). The use of facial motion and facial form during the processing of identity. *Vision Research*, *43*, 1921–1936.

LaBar, K. S., Crupain, M. J., Voyvodic, J. T., & McCarthy, G. (2003). Dynamic perception of facial affect and identity in the human brain. *Cerebral Cortex*, *13*, 1023–1033.

Lansing, C. R., & McConkie, G. W. (2003). Word identification and eye fixation locations in visual and visual-plus-auditory presentations of spoken sentences. *Attention, Perception, & Psychophysics*, *65*, 536–552.

Lawrence, K., Bernstein, D., Pearson, R., Mandy, W., Campbell, R., & Skuse, D. (2008). Changing abilities in recognition of unfamiliar face photographs through childhood and adolescence: Performance on a test of non-verbal immediate memory (Warrington RMF) from 6 to 16 years. *Journal of Neuropsychology*, *2*, 27–45.

Lawrence, K., Campbell, R., & Skuse, D. (2015). Age, gender, and puberty influence the development of facial emotion recognition. *Frontiers in Psychology*, *6*, 761.

Liu, J., Harris, A., & Kanwisher, N. (2002). Stages of processing in face perception: An MEG study. *Nature Neuroscience*, *5*, 910–916.

Liu, J., Harris, A., & Kanwisher, N. (2010). Perception of face parts and face configurations: An fMRI study. *Journal of Cognitive Neuroscience*, *22*, 203–211.

Lord, C., Risi, S., Lambrecht, L., Cook, E. H., Leventhal, B. L., DiLavore, P. C., … Rutter, M. (2000). The autism diagnostic observation schedule—Generic: A standard measure of social and communication deficits associated with the spectrum of autism. *Journal of Autism and Developmental Disorders*, *30*, 205–223.

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

MacLeod, A., & Summerfield, Q. (1987). Quantifying the contribution of vision to speech perception in noise. *British Journal of Audiology*, *21*, 131–141.

Maguinness, C., Setti, A., Burke, K., Kenny, R. A., & Newell, F. N. (2011). The effect of combined sensory and semantic components on audio-visual speech perception in older adults. *Frontiers in Aging Neuroscience*, *3*, 19.

Marassa, L. K., & Lansing, C. R. (1995). Visual word recognition in two facial motion conditions: Full-face versus lips-plus-mandible. *Journal of Speech, Language, and Hearing Research*, *38*, 1387–1394.

Nichols, T., & Hayasaka, S. (2003). Controlling the familywise error rate in functional neuroimaging: A comparative review. *Statistical Methods in Medical Research*, *12*, 419–446.

O'Brien, J., Spencer, J., Girges, C., Johnston, A., & Hill, H. (2014). Impaired perception of facial motion in autism spectrum disorder. *PLoS One*, *9*, e102173.

O'Reilly, J. X., Woolrich, M. W., Behrens, T. E., Smith, S. M., & Johansen-Berg, H. (2012). Tools of the trade: Psychophysiological interactions and functional connectivity. *Social Cognitive and Affective Neuroscience*, *7*, 604–609.

Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia*, *9*, 97–113.

Omidikia, N., & Kompany-Zareh, M. (2013). Uninformative variable elimination assisted by gram–Schmidt orthogonalization/successive projection algorithm for descriptor selection in QSAR. *Chemometrics and Intelligent Laboratory Systems*, *128*, 56–65.

OToole, A. J., Roark, D. A., & Abdi, H. (2002). Recognizing moving faces: A psychological and neural synthesis. *Trends in Cognitive Sciences*, 6, 261–266.

Palermo, R., Willis, M. L., Rivolta, D., McKone, E., Wilson, C. E., & Calder, A. J. (2011). Impaired holistic coding of facial expression and facial identity in congenital prosopagnosia. *Neuropsychologia*, 49, 1226–1235.

Pelphrey, K. A., Morris, J. P., McCarthy, G., & LaBar, K. S. (2007). Perception of dynamic changes in facial affect and identity in autism. *Social Cognitive and Affective Neuroscience*, 2, 140–149.

Pitcher, D., Duchaine, B., & Walsh, V. (2014). Combined TMS and fMRI reveal dissociable cortical pathways for dynamic and static face perception. *Current Biology*, 24, 2066–2070.

Pitcher, D., Walsh, V., & Duchaine, B. (2011). The role of the occipital face area in the cortical face perception network. *Experimental Brain Research*, 209, 481–493.

Pitcher, D., Walsh, V., Yovel, G., & Duchaine, B. (2007). TMS evidence for the involvement of the right occipital face area in early face processing. *Current Biology*, 17, 1568–1573.

Puce, A., Allison, T., Bentin, S., Gore, J. C., & McCarthy, G. (1998). Temporal cortex activation in humans viewing eye and mouth movements. *Journal of Neuroscience*, 18, 2188–2199.

Pyles, J. A., Verstynen, T. D., Schneider, W., & Tarr, M. J. (2013). Explicating the face perception network with white matter connectivity. *PLoS One*, 8, e61611.

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.

Rosenblum, L. D., Johnson, J. A., & Saldaña, H. M. (1996). Point-light facial displays enhance comprehension of speech in noise. *Journal of Speech, Language, and Hearing Research*, 39, 1159–1170.

Rosenblum, L. D., & Saldaña, H. M. (1996). An audiovisual test of kinematic primitives for visual speech perception. *Journal of Experimental Psychology-Human Perception and Performance*, 22, 318–330.

Rosenblum, L. D., & Saldaña, H. M. (1998). Time-varying information for visual speech perception. In R. Campbell, B. Dodd, & D. Burnham (Eds.), *Hearing by Eye II: Advances in the psychology of speechreading and auditory-visual speech* (pp. 61–81). Hove, UK: Psychology Press, Ltd.

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. *Cerebral Cortex*, 17, 1147–1153.

Rossion, B. (2008). Constraining the cortical face network by neuroimaging studies of acquired prosopagnosia. *NeuroImage*, 40, 423–426.

Rossion, B., Hanseeuw, B., & Dricot, L. (2012). Defining face perception areas in the human brain: A large-scale factorial fMRI face localizer analysis. *Brain and Cognition*, 79, 138–157.

Rouger, J., Lagleyre, S., Fraysse, B., Deneve, S., Deguine, O., & Barone, P. (2007). Evidence that cochlear-implanted deaf patients are better multisensory integrators. *Proceedings of the National Academy of Sciences*, 104, 7295–7300.

Rühl, D., Bölte, S., Feineis-Matthews, S., & Poustka, F. A. D. O. S. (2004). *Diagnostische Beobachtungsskala für Autistische Störungen (ADOS)*. Bern: Huber.

Ruxton, G. D. (2006). The unequal variance t-test is an underused alternative to Student's t-test and the Mann–Whitney U test. *Behavioral Ecology*, 17, 688–690.

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. *Journal of Autism and Developmental Disorders*, 42, 1606–1615.

Sabatinelli, D., Fortune, E. E., Li, Q., Siddiqui, A., Krafft, C., Oliver, W. T., ... Jeffries, J. (2011). Emotional perception: Meta-analyses of face and natural scene processing. *NeuroImage*, 54, 2524–2533.

Sato, W., Kochiyama, T., Yoshikawa, S., Naito, E., & Matsumura, M. (2004). Enhanced neural activity in response to dynamic facial expressions of emotion: An fMRI study. *Cognitive Brain Research*, 20, 81–91.

Sato, W., Toichi, M., Uono, S., & Kochiyama, T. (2012). Impaired social brain network for processing dynamic facial expressions in autism spectrum disorders. *BMC Neuroscience*, 13, 99.

Sato, W., Uono, S., & Toichi, M. (2013). Atypical recognition of dynamic changes in facial expressions in autism spectrum disorders. *Research in Autism Spectrum Disorders*, 7, 906–912.

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.

Schultz, J., & Pilz, K. S. (2009). Natural facial motion enhances cortical responses to faces. *Experimental Brain Research*, 194, 465–475.

Schweinberger, S. R., & Soukup, G. R. (1998). Asymmetric relationships among perceptions of facial identity, emotion, and visual speech. *Journal of Experimental Psychology: Human Perception and Performance*, 24, 1748.

Smith, E. G., & Bennetto, L. (2007). Audiovisual speech integration and lipreading in autism. *Journal of Child Psychology and Psychiatry*, 48, 813–821.

Smith, S. M., Jenkinson, M., Woolrich, M. W., Beckmann, C. F., Behrens, T. E., Johansen-Berg, H., ... Niazy, R. K. (2004). Advances in functional and structural MR image analysis and implementation as FSL. *NeuroImage*, 23, 208–219.

Stevenson, R. A., Siemann, J. K., Schneider, B. C., Eberly, H. E., Woynaroski, T. G., Camarata, S. M., & Wallace, M. T. (2014). Multisensory temporal integration in autism spectrum disorders. *Journal of Neuroscience*, 34, 691–697.

Stone, L. (1957). *Facial clues of context in lip reading*. Los Angeles: John Tracy Clinic.

Sumby, W. H., & Pollack, I. (1954). Visual contributions to speech intelligibility in noise. *Journal of the Acoustical Society of America*, 26, 212–215.

Tanaka, J. W., Kaiser, M. D., Butler, S., & Le Grand, R. (2012). Mixed emotions: Holistic and analytic perception of facial expressions. *Cognition & Emotion*, 26, 961–977.

Thomas, S. M., & Jordan, T. R. (2002). Determining the influence of Gaussian blurring on inversion effects with talking faces. *Perception & Psychophysics*, 64, 932–944.

Thomas, S. M., & Jordan, T. R. (2004). Contributions of oral and extraoral facial movement to visual and audiovisual speech perception. *Journal of Experimental Psychology: Human Perception and Performance*, 30, 873.

Travers, B. G., Adluru, N., Ennis, C., Tromp, D. P., Destiche, D., Doran, S., ... Alexander, A. L. (2012). Diffusion tensor imaging in autism spectrum disorder: A review. *Autism Research*, 5, 289–313.

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

Von Aster, M., Neubauer, A., & Horn, R. (2006). *Wechsler Intelligenztest Für Erwachsene (WIE)*. Frankfurt/M: Harcourt Test Services.

Von Kriegstein, K., Dogan, Ö., Grüter, M., Giraud, A. L., Kell, C. A., Grüter, T., ... Kiebel, S. J. (2008). Simulation of talking faces in the human brain improves auditory speech recognition. *Proceedings of the National Academy of Sciences*, 105, 6747–6752.

Von Kriegstein, K., & Giraud, A. L. (2004). Distinct functional substrates along the right superior temporal sulcus for the processing of voices. *NeuroImage*, 22, 948–955.

Vuilleumier, P., Armony, J. L., Driver, J., & Dolan, R. J. (2001). Effects of attention and emotion on face processing in the human brain: An event-related fMRI study. *Neuron*, 30, 829–841.

Wakabayashi, A., Baron-Cohen, S., Wheelwright, S., & Tojo, Y. (2006). The Autism-Spectrum Quotient (AQ) in Japan: A cross-cultural comparison. *Journal of Autism and Developmental Disorders*, 36, 263–270.

Wechsler, D. (1997). *Wechsler Adult Intelligence Scale (WAIS-III)*. San Antonio, TX: The Psychological Corporation.

Weigelt, S., Koldewyn, K., & Kanwisher, N. (2012). Face identity recognition in autism spectrum disorders: A review of behavioral studies. *Neuroscience & Biobehavioral Reviews*, *36*, 1060–1084.

World Health Organization. (2004). *International statistical classification of diseases and related and related health problems* (10th ed.). Geneva: World Health Organization.

Wright, T. M., Pelphrey, K. A., Allison, T., McKeown, M. J., & McCarthy, G. (2003). Polysensory interactions along lateral temporal regions evoked by audiovisual speech. *Cerebral Cortex*, *13*, 1034–1043.

Yovel, G., Tambini, A., & Brandman, T. (2008). The asymmetry of the fusiform face area is a stable individual characteristic that underlies the left-visual-field superiority for faces. *Neuropsychologia*, *46*, 3061–3068.

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. *Journal of Neuroscience, 11*, 641–649.

Zhu, Q., Song, Y., Hu, S., Li, X., Tian, M., Zhen, Z., ... Liu, J. (2010). Heritability of the specific cognitive ability of face perception. *Current Biology*, *20*, 137–142.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.