

The relationship between level of autistic traits and local bias in the context of the McGurk effect

Yuta Ujiie^{1,2*}, Tomohisa Asai³ and Akio Wakabayashi⁴

¹ Information Processing and Computer Sciences, Graduate School of Advanced Integration Science, Chiba University, Chiba, Japan, ² Japan Society for the Promotion of Science, Tokyo, Japan, ³ NTT Communication Science Laboratories, NTT Corporation, Kanagawa, Japan, ⁴ Faculty of Letters, Chiba University, Chiba, Japan

OPEN ACCESS

Edited by:

Snehlata Jaswal,
Indian Institute of Technology,
Jodhpur, India

Reviewed by:

Ankita Sharma,
Indian Institute of Technology,
Jodhpur, India
Kaisa Tiippana,
University of Helsinki, Finland

*Correspondence:

Yuta Ujiie,
Information Processing and Computer
Sciences, Graduate School of
Advanced Integration Science, Chiba
University, 1-33 Yayoi-cho, Inage,
Chiba 263-8522, Japan
chiba_psyg_individual@yahoo.co.jp

Specialty section:

This article was submitted to
Cognitive Science,
a section of the journal
Frontiers in Psychology

Received: 15 March 2015

Accepted: 15 June 2015

Published: 30 June 2015

Citation:

Ujiie Y, Asai T and Wakabayashi A
(2015) The relationship between level
of autistic traits and local bias in the
context of the McGurk effect.
Front. Psychol. 6:891.
doi: 10.3389/fpsyg.2015.00891

The McGurk effect is a well-known illustration that demonstrates the influence of visual information on hearing in the context of speech perception. Some studies have reported that individuals with autism spectrum disorder (ASD) display abnormal processing of audio-visual speech integration, while other studies showed contradictory results. Based on the dimensional model of ASD, we administered two analog studies to examine the link between level of autistic traits, as assessed by the Autism Spectrum Quotient (AQ), and the McGurk effect among a sample of university students. In the first experiment, we found that autistic traits correlated negatively with fused (McGurk) responses. Then, we manipulated presentation types of visual stimuli to examine whether the local bias toward visual speech cues modulated individual differences in the McGurk effect. The presentation included four types of visual images, comprising no image, mouth only, mouth and eyes, and full face. The results revealed that global facial information facilitates the influence of visual speech cues on McGurk stimuli. Moreover, individual differences between groups with low and high levels of autistic traits appeared when the full-face visual speech cue with an incongruent voice condition was presented. These results suggest that individual differences in the McGurk effect might be due to a weak ability to process global facial information in individuals with high levels of autistic traits.

Keywords: autism spectrum disorder, Autism Spectrum Quotient, the McGurk effect, local bias, individual differences

Introduction

Autism spectrum disorder (ASD) has been largely defined in terms of difficulties in social interaction and communication, patterns of repetitive behavior, and narrow interests (American Psychiatric Association, 1994, 2013). In earlier ASD research, the dysfunction of processing information relevant to social interaction was the main focus of investigations. It has been revealed that individuals with ASD show different patterns in face perception (e.g., Deruelle et al., 2004) and emotion recognition (e.g., Baron-Cohen et al., 2001a) compared to individuals with typical development (TD). In addition to dysfunction in processing visual stimuli, recent studies have shown that individuals with ASD exhibit atypical processing in audio-visual speech perception (Massaro and Bosseler, 2006; Smith and Bennetto, 2007), which indicates a limited ability to integrate visual and auditory information. This dysfunction is considered to lead to communication impairment in ASD because speech perception is one of the core functions of face-to-face communication.

In face-to-face communication with others, we realize what another person is saying through the processing of audio-visual speech information. An early study of audio-visual speech perception provided strong evidence that visual information improves the auditory speech percept (Sumbly and Pollack, 1954). A classic example that demonstrates the interaction between hearing and vision in speech perception is the McGurk effect (McGurk and MacDonald, 1976). This effect may be experienced when the visual shape produced during speech of a phoneme (e.g., /ga/) is dubbed with a sound recording of a different phoneme (e.g., /ba/), which often causes a third, intermediate phoneme (e.g., /da/) to be perceived. Similarly, for the monosyllabic combination of the visual /ka/ and auditory /pa/, participants often reported hearing /ta/.

Abnormal processing of audio-visual speech integration in individuals with ASD has been reported in a number of studies (De Gelder et al., 1991; Williams et al., 2004; Iarocci et al., 2010; Taylor et al., 2010; Saalasti et al., 2011, 2012; Woynaroski et al., 2013; Stevenson et al., 2014). Iarocci et al. (2010) reported that children with ASD showed less visual influence and more auditory influence during bimodal speech perception than controls did, due to poor lip-reading ability, and this finding was supported by Williams et al. (2004). On the other hand, De Gelder et al. (1991) reported that children with ASD are less influenced by the auditory percepts from visual speech cues, although they did not differ in lip-reading ability from children with TD. Similar results have been reported in children with ASD (Stevenson et al., 2014) and in adults with ASD (Saalasti et al., 2011, 2012). Although, the results are mixed, it has often been reported that individuals with ASD exhibit a weak degree of visual influence on perceiving a voice during audio-visual speech perception.

Another study suggested that the reduced McGurk effect in individuals with ASD meant that there was a delay, rather than a deficit, in the development of audio-visual integration (Taylor et al., 2010). Taylor et al. (2010) showed that younger children with ASD exhibit delayed visual accuracy and audio-visual integration (the McGurk effect) compared to children with TD, but appeared to catch up with their TD peers in the older age ranges. In line with this, Keane et al. (2010) revealed no individual differences in the McGurk effect between adults with and without ASD. This was inconsistent with the results in some previous studies (Saalasti et al., 2012; Stevenson et al., 2014).

Neuropsychological data provide us with some advantages to identify whether individuals with ASD show weaker visual influence on perceiving a voice. In ASD, several studies have reported anatomical and functional abnormalities in the superior temporal sulcus (STS) (see, for a review, Zilbovicius et al., 2006; Redcay, 2008). The STS is critical for integration of auditory and visual speech information, and influences the likelihood of the McGurk effect occurring (Calvert et al., 2000; Nath and Beauchamp, 2011). Redcay (2008) argued that impairments in STS function might lead to abnormalities in speech perception in individuals with ASD. In the sample of individuals with TD, a functional magnetic resonance imaging (fMRI) study revealed a significant positive correlation between the likelihood of perceiving the McGurk effect and the amplitude of the

response in the left STS (Nath and Beauchamp, 2012). This means that individuals with a weak response in the left STS showed fewer instances of the McGurk effect when they observed audio-visual-incongruent stimuli. Another study using near-infrared spectroscopy (NIRS) reported a significantly negative correlation between level of autistic traits and regional cerebral blood volume in the left STS during face-to-face conversation among adults with TD (Suda et al., 2011). These studies led us to hypothesize that individuals with ASD (or a high level of autistic traits) might show fewer instances of the McGurk effect due to a weak response in the left STS.

The reason why previous results are mixed might be due to the heterogeneity of the clinical population. For instance, abnormalities in sensory inputs (hyper- or hypo-sensitivity), which is one of the core symptoms of ASD, have been found in more than 90% of individuals with ASD in at least one sensory domain (Tomchek and Dunn, 2007; Crane et al., 2009), but in which sensory domain the abnormality appears varies (e.g., visual: Simmons et al., 2009; auditory: Haesen et al., 2011; O'Connor, 2012; tactile: Foss-Feig et al., 2012). With regard to the McGurk effect, one study showed that a fusion response was correlated with the degree of auditory processing difficulty, as assessed by the Sensory Profile (Dunn and Westman, 1997) among individuals with ASD (Woynaroski et al., 2013). Saalasti et al. (2012) examined the distribution of the likelihood of the McGurk response occurring and showed that the difference between the control group and the clinical group with ASD was significant. Mixed results in previous studies might be due to the heterogeneity in the profile of hyper- or hypo-sensitivity, which is difficult to control in a clinical group.

An analog design is one approach used to study ASD symptoms among individuals with TD, by examining the relationship between level of autistic traits and performance on a cognitive or perceptual task. This approach is based on the dimensional model of ASD, which assumes that autistic traits are distributed on a continuum over clinical and general populations (Frith, 1991; Baron-Cohen, 1995). In order to assess the degree of autistic traits in any individual adult with a normal intelligence quotient, Baron-Cohen et al. (2001b) developed the Autism Spectrum Quotient (AQ). The AQ is a self-report questionnaire and is useful as a screening scale to not only distinguish between clinical and control groups but also measure the distribution of autistic traits within the general population. The validity and reliability of this screening scale have been confirmed in various countries (UK: Baron-Cohen et al., 2001b; the Netherlands: Hoekstra et al., 2008; Australia: Lau et al., 2013; and Japan: Wakabayashi et al., 2006b). Moreover, our pilot study (Ujiié and Wakabayashi, 2015) found that the overlap between level of autistic traits and the degree of hyper- or hypo-sensitivity, which was assessed by the Glasgow Sensory Questionnaire (Robertson and Simmons, 2012), was small among a general population. Because an analog design allows for control of the heterogeneity in the profile of hyper- or hypo-sensitivity, we adopted this design to examine the relationship between level of autistic traits and McGurk effects among a population with TD who were free from problems with sensory inputs.

The purpose of this study was to use an analog design based on the dimensional model of ASD to investigate the relationship between individual differences in the McGurk effect and autistic traits in the general population. First, we investigated whether autistic traits were correlated with a weaker visual influence on speech perception under a without-noise condition (Experiment 1). As the McGurk stimuli, we used the combination of auditory /pa/ and visual /ka/ stimuli, because this combination was likely to be perceived as a stronger illusion than other combinations (e.g., auditory /ba/ and visual /ga/ stimuli) in a Japanese sample (Sekiyama, 1994). There were three possible responses to the McGurk stimuli, which were as follows: audio response (/pa/ response), fused response (/ta/ response), and visual response (/ka/ response). In this study, we defined the rate of fused response, which was the frequency of the McGurk effect occurring, as the degree of visually captured percept when hearing and viewing the McGurk stimuli. We defined the rate of /pa/ response, which was the correct response to the audio-visual-incongruent stimuli, as the strength of visual influence on perceiving a voice. Thus, we hypothesized that the degree of autistic traits would correlate negatively with the rate of fused response and correlate positively with the rate of audio response in the context of the McGurk stimuli. In Experiment 2, we focused on the local bias toward visual speech cues in individuals with ASD, and investigated whether this bias underlies the link between level of autistic traits and the McGurk effect.

Individuals with ASD have been shown to tend to prefer local over global information when presented multiple information sources (Happe and Frith, 2006). Such cognitive specificity in individuals with ASD is called the local bias (Frith, 1991; Happe and Frith, 2006), and this bias has been mainly reported in relation to visuospatial tasks. For instance, individuals with ASD show better performance than individuals with TD on the Embedded Figures Test (EFT; Brosnan et al., 2012), in which one is required to find a local (embedded) target within a global context constructed of multiple figures. A similar result has been found in the Navon-type Global–Local Naming Task (Reed et al., 2011) and various perceptual and cognitive tasks (Happe and Frith, 2006).

Furthermore, the local bias has been found in face processing tasks (Joseph and Tanaka, 2003; Deruelle et al., 2004; Kätsyri et al., 2008). Deruelle et al. (2004) investigated whether children with ASD preferred to use local (high spatial frequency) rather than global (low spatial frequency) information during a face matching task. The results showed that children with ASD showed better performance when using local information than when they used global information. Kätsyri et al. (2008) found a similar result in the recognition of dynamic facial emotions among adults with ASD. In addition, some studies suggested a preference of gaze toward the mouth region when individuals with ASD perceived face stimuli (Klin et al., 2002; Joseph and Tanaka, 2003). Joseph and Tanaka (2003) revealed that individuals with ASD use more information from around the mouth region of face stimuli in face recognition tasks. Klin et al. (2002) showed that individuals with ASD tend to gaze at the mouth region during the viewing of conversation. These results indicate that individuals with ASD might prefer to use local information, particularly that from the mouth region, during

face processing. Individuals with high AQ scores who exhibit a local bias (Reed et al., 2011) might be less likely to experience the McGurk effect. To clarify this, we should examine whether face processing influences the occurrence of the McGurk effect, in general.

Whether face processing is needed for audio-visual speech perception has been discussed in the past. According to Bruce and Young's (1986) model, face perception has three important functions, comprising recognition of facial identity, facial expression, and facial speech. In line with this suggestion, some studies have demonstrated a relationship between face processing and speech perception. De Gelder et al. (1991) showed that face identification correlates positively with the influence of lip-reading on audio-visual-incongruent stimuli. Rosenblum et al. (2002) showed that face processing and speech perception share the same dynamic information. Some studies, however, suggested that a mouth-only presentation influenced voice perception and produced the McGurk effect as well as a whole-face presentation did (Rosenblum et al., 2000; Hietanen et al., 2001). Other studies showed the role of extraoral facial information in audio-visual speech perception (Thomas and Jordan, 2004; Jordan and Thomas, 2011). Jordan and Thomas (2011) revealed that occluded oral areas disrupted performance but that observers could use lip-reading and observe visual speech influences from extraoral areas. Thomas and Jordan (2004) showed that an extraoral movement during visual speech was effective to perceive visual speech cues and influenced audio-visual speech perception.

The role of holistic face processing on speech perception has been investigated by using methods to examine the effect of holistic processing on face perception. For instance, some studies investigated the face inversion effect, which is a phenomenon that causes difficulty in holistic face processing by showing an inverted face, in the context of audio-visual speech perception (Jordan and Bevan, 1997; Rosenblum et al., 2000; Eskelund et al., 2015). One study showed a robust effect (Eskelund et al., 2015) while another study found only a partial effect (Jordan and Bevan, 1997), which means that the face inversion effect depends on the stimulus. Rosenblum et al. (2000) examined the role of holistic facial information in audio-visual speech perception, compared to full-face Thatcher-type speech stimuli and inverted mouth-alone speech stimuli. In their study, the full-face Thatcher-type stimuli were created by combining an upright face with an inverted mouth visual speech cue. They reported that only the full-face Thatcher-type stimuli disrupted voice perception for the audio-visual-incongruent condition (the combination of auditory /va/ and visual /ba/). This result, which is called the McThatcher effect, was replicated in Eskelund et al. (2015). Hietanen et al. (2001) investigated the effect of facial configuration context on the McGurk effect. They manipulated the location of facial features in visual stimuli, using either a natural or scrambled location. In their results, only an asymmetrically scrambled face disrupted the likelihood of the McGurk effect, but this effect depended on the stimulus. They concluded that facial configuration information can be used in audio-visual speech perception, although this information is not necessary. These studies indicate that processing of global

(holistic) visual speech cues might influence the occurrence of the McGurk effect.

In summary, based on the dimensional model of ASD, we administered two experiments to examine the relationship between level of autistic traits and McGurk effects in university students. In Experiment 1, we investigated the correlation between level of autistic traits and individual differences in audio-visual speech perception. We hypothesized that individuals with high levels of autistic traits would show a reduced likelihood of the McGurk effect occurring than would individuals with low levels of autistic traits. In Experiment 2, we examined whether the local bias toward visual speech cues modulates individual differences in the McGurk effect, by manipulating presentation types of visual stimuli (parts of the face or the full face). With regard to the likelihood of the McGurk effect occurring, we hypothesized that the visual influence on voice perception would be greater in the full-face presentation condition than in the partial-face presentation condition. In addition, we hypothesized that individual differences in the McGurk effect would appear when the full-face visual speech cue with an incongruent voice condition was presented, because of the local bias toward visual speech cues in individuals with high levels of autistic traits. The outcomes from these experiments will allow us to understand the effect of face processing on speech perception, and how audio-visual speech integration in individuals with ASD functions from an analog perspective.

Experiment 1

In Experiment 1, we investigated the correlation between AQ scores and level of accuracy for perceiving audio-visual stimuli and auditory stimuli, and assessed the likelihood of the McGurk effect (the rate of /ta/ response) occurring among non-ASD university students. For audio-visual-incongruent stimuli, we hypothesized that, because of the weak visual influence of perceiving a voice, the AQ scores would correlate negatively with the likelihood of the McGurk effect occurring, and positively with the rate of the /pa/ response being reported.

Methods

Participants

Participants were 46 university students (12 males and 34 females) who were recruited from an introductory psychology class at Chiba University. The mean age of the participants was 19.4 years ($SD = 3.56$). All participants were native speakers of Japanese and reported normal hearing and vision. They provided written informed consent in the class, and took part voluntarily in this experiment. After the experiment, they received an oral debriefing.

Stimuli

Japanese version of the AQ

The AQ was normalized for use in the Japanese population by Wakabayashi et al. (2006b). The AQ contains 50 items for assessing the following five domains: social skill, attention switching, attention to detail, communication, and imagination. Participants rate each item on a 4-point response scale from

“agree” to “disagree.” Each item is scored 0 or 1 point according to the scoring manner described in previous studies (Baron-Cohen et al., 2001b; Wakabayashi et al., 2006b), so that total scores on the AQ range from 0 to 50.

Audio-visual task

The audio-visual stimuli were created from simultaneous audio and video recordings of six Japanese speakers' utterances (three female). The visual stimuli were speakers' faces recorded using a digital video camera (GZ-EX370, JVC KENWOOD). The audio stimuli were the utterances (/pa/, /ta/, or /ka/) collected using a dynamic microphone (MD42, SENNHEISER). The video clip (720 × 480 pixels, 29.97 frames/s) and the speech sound (digitized at 48,000 Hz, with 16-bit quantization resolution) were combined and synchronized using Adobe Premiere Pro CS6. The mean duration of the audio-visual stimuli was 1.2 s.

There were three stimulus conditions, comprising audio-only (e.g., auditory /pa/), audio-visual congruent (e.g., auditory /pa/, visual /pa/), and audio-visual incongruent (e.g., auditory /pa/, visual /ka/). Each condition included 18 trials per block.

In the audio-only condition, the audio stimuli (/pa/, /ta/, or /ka/) were presented without the visual stimuli. In the audio-visual-congruent condition, all three combinations of the audio and visual stimuli were presented. In the audio-visual-incongruent condition, the combination of the auditory /ka/ stimulus dubbed with visual /pa/ was excluded, because the percept (e.g., /pka/) caused by this combination stimulus is not a Japanese native syllable. Therefore, the voice (/pa/) and video (/ka/) combined stimuli were presented three times per block to make the same number of audio-visual-congruent trials.

Apparatus

The experiment was conducted using Hot Soup Processor Version 3.3 (Onion software). The video signals were presented on a 19-inch cathode ray tube (CRT) monitor (E193FPp, Dell), and the speech sound was presented through a headphone (MDR-Z500, Sony) at approximately a 65 dB sound pressure level, adjusted using a mixing console (MW8CX, Yamaha).

Procedure

Participants were seated at a distance of approximately 50 cm from the CRT monitor, wearing the headphone. Participants were instructed to report what they heard (/pa/, /ta/, or /ka/) by a key press. In each trial, a fixation point was displayed for 1000 ms at the center of the CRT monitor, followed by either the congruent or the incongruent stimulus. Then, a blank display was presented until participants responded.

The first block included 18 congruent stimuli and 18 incongruent stimuli. The second block included 18 auditory stimuli. All participants completed both blocks after undergoing six practice trials each. The order of trials was randomized for each block. After all of the tasks were finished, participants completed the questionnaire.

Data Analysis

Statistical analysis was conducted using R version 2.15.2 for Windows (R Foundation for Statistical Computing, Vienna,

Austria). To examine the effect of stimuli conditions, we analyzed the mean accuracies for each condition using a One-Way analysis of variance (ANOVA), with conditions as a within-participants factor. The likelihood of the McGurk effect occurring was analyzed in the audio-visual-incongruent condition using a chi-square test. The relationship between task performance and AQ scores was analyzed using Pearson correlation coefficients. In addition, group differences between a high-AQ group and a low-AQ group were analyzed using independent samples *t*-tests for each condition.

Results

Table 1 shows mean accuracies for the audio-visual-congruent condition and the audio-only condition, and the mean response rate for the audio-visual-incongruent condition. A One-Way ANOVA with conditions as a within-participants factor revealed a main effect of conditions, $F_{(2, 90)} = 197.215$, $p < 0.01$, partial $\eta^2 = 0.81$. Multiple comparisons (Holm method) showed that accuracies for correctly perceiving the voice in the audio-only condition ($M = 97.3\%$) and the audio-visual-congruent condition ($M = 98.1\%$) were higher than in the audio-visual-incongruent condition ($M = 34.9\%$; $p < 0.05$). However, the accuracy in the audio-only condition did not differ from that in the audio-visual-congruent condition. In the audio-visual-incongruent condition, the rate of the /ta/ response ($M = 61.1\%$) was higher than the rate of the /pa/ response ($M = 34.9\%$; $\chi^2 = 7.66$, $p < 0.01$) and the /ka/ response ($M = 4.0\%$; $\chi^2 = 20.33$, $p < 0.01$), which confirmed the occurrence of the McGurk effect.

The AQ scores ranged from 10 to 37 ($M = 20.8$, $SD = 5.42$). The distribution of AQ scores in this sample was slightly higher than that reported in the original publication of the AQ (Baron-Cohen et al., 2001b). To examine the relationship between task performances and the AQ scores, we calculated Pearson correlation coefficients (**Table 1**). No significant correlation was observed for the audio-visual-congruent condition or the audio-only condition. For the audio-visual-incongruent condition, the AQ was significantly positively correlated with the /pa/ response,

$r_{(46)} = 0.29$, $p < 0.05$, and significantly negatively correlated with the /ta/ response, $r_{(46)} = -0.32$, $p < 0.05$. These correlations suggest that individuals with low AQ scores show a more visually captured response and less audio response than individuals with high AQ scores do.

Next, we examined group differences between individuals with high AQ scores and those with low AQ scores in each condition. From among the participants, we picked eight with scores of 15 or under (mean AQ $- 1$ SD), and another eight with scores of 26 or over (mean AQ $+ 1$ SD). We regarded the former as the low-AQ group (4 males and 4 females, mean AQ = 13.5) and the latter as the high-AQ group (3 males and 5 females, mean AQ = 29.3). A between-groups *t*-test showed a significant difference in the AQ scores, $t_{(14)} = 11.26$, $p < 0.01$, $r = 0.95$. Similarly, we conducted independent samples *t*-tests for each condition. No significant difference was found in the audio-only condition, $t_{(14)} = 1.17$, *ns*, $r = 0.29$, or in the audio-visual-congruent condition, $t_{(14)} = 0.43$, *ns*, $r = 0.12$ (see Supplementary Material). For the audio-visual-incongruent condition (see **Figure 1**), the rate of the /ta/ response was higher in the low-AQ group ($M = 65.3\%$) than in the high-AQ group ($M = 43.1\%$). This difference was marginally significant, $t_{(14)} = 1.79$, $p < 0.10$, $r = 0.43$; however, the rate of the /pa/ response was not significantly different, $t_{(14)} = 1.62$, $p = 0.12$, $r = 0.40$. These results indicate that individuals with high AQ scores show weaker visually captured responses than individuals with low AQ scores do, although accuracies for perceiving voice and audio-visual speech did not differ.

Discussion

In this experiment, we investigated the relationship between audio-visual speech integration and the level of autistic traits in healthy students. We found that the level of autistic traits correlated negatively with the rate of fused response and positively with the rate of audio response in the audio-visual-incongruent condition. Moreover, the results revealed that individuals with high AQ scores showed a weaker fused response

TABLE 1 | Mean and standard deviations (SD) of response rate in all conditions.

	Mean	SD	Correlations with the AQ scores
AUDIO-ONLY CONDITION			
Correct response (/pa/, /ta/, /ka/)	0.97	0.03	0.00
AUDIO-VISUAL-CONGRUENT CONDITION			
Correct response (/pa/, /ta/, /ka/)	0.99	0.03	-0.10
AUDIO-VISUAL-INCONGRUENT CONDITION			
Audio response (/pa/)	0.35	0.29	0.29*
Fused response (/ta/)	0.61	0.28	-0.30*
Visual response (/ka/)	0.04	0.08	-0.19

Also displayed are correlations between the AQ scores and response rates for experiment 1. $N = 46$. AQ, Autism-spectrum Quotient. Correct response (/pa/, /ta/, /ka/): Mean correct responses for all stimuli in the audio-visual-congruent condition or the audio-only condition. * $p < 0.05$, ** $p < 0.01$.

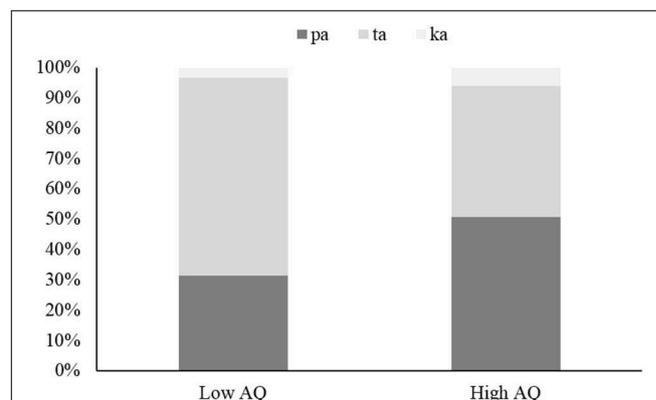


FIGURE 1 | The response rate for each audio-visual-incongruent stimulus in the low-AQ group and the high-AQ group. Possible responses to the stimuli were audio response (/pa/ response), fused response (/ta/ response), and visual response (/ka/ response).

than individuals with low AQ scores did, although there was no significant difference in the audio response rate. On the other hand, neither significant correlations nor group differences were found in the audio-visual-congruent condition and audio-only condition. These results indicate that individuals with higher levels of autistic traits tended to show a weaker visual influence on perceiving a voice when processing audio-visual-incongruent speech information.

Several studies reported that individuals with ASD showed a weaker visual influence only when McGurk stimuli are presented (e.g., De Gelder et al., 1991). This study replicated those results in a sample of university students. As we hypothesized, our results indicate that the weakness of visual influence on audio-visual speech perception exists along the distribution of AQ in the general population. This finding might support the dimensional model of ASD, because individuals with high AQ scores in this study and individuals with ASD in previous studies (e.g., De Gelder et al., 1991) showed a similar tendency when processing audio-visual-incongruent speech.

However, in this experimental task, it was not clear what factor led to a weaker visual influence on audio-visual speech perception. One possibility is that the local bias toward visual speech cues reflected individual differences in the McGurk effect. There was a local bias effect of cognitive specificity on individuals with ASD, meaning that there is a bias toward processing local information in preference to global information (Frith, 1991; Happe and Frith, 2006). In addition to the results for visuo-spatial tasks (Reed et al., 2011; Brosnan et al., 2012), recent studies have reported that individuals with ASD have a preference for feature-based processing of face stimuli (Joseph and Tanaka, 2003; Deruelle et al., 2004), and for focusing on the local (mouth) region during the viewing of conversation videos (Klin et al., 2002). Some studies have suggested that the influence of visual speech cues in processing audio-visual-incongruent stimuli is related to the processing of faces (De Gelder et al., 1991), especially in the global (holistic) facial context (Rosenblum et al., 2000). Thus, if global facial context enhances the influence of visual speech cues on perceiving a voice, individual differences in the McGurk effect between individuals with high AQ scores and those with low AQ scores might be due to a weak ability to process global facial information in McGurk stimuli. To confirm this, we conducted Experiment 2.

Experiment 2

In Experiment 2, we manipulated presentation types of visual stimuli to examine whether the local bias affected individual differences in the McGurk effect. We set two stimulus conditions, i.e., the audio-visual-congruent condition and the audio-visual-incongruent condition. For the audio-visual-incongruent condition, we defined the rate of fused responses as the frequency of visually captured percept, while we defined the rate of /pa/ responses as the strength of visual influence to perceiving voice, as in Experiment 1.

In addition, we created the following four types of visual stimuli: no image (audio-only), mouth-only, eyes and mouth, and full face. Only the full-face stimuli included global

facial information of visual speech cues. For the audio-visual-incongruent condition, we hypothesized that audio response would be observed less frequently in the full-face presentation than in the other stimuli conditions if the processing of global visual speech cues is related to the degree of visual influence on perceiving a voice. Moreover, we also hypothesized that the differences between individuals with high AQ scores and those with low AQ scores would diminish (or become small) when a voice and an incongruent visual speech cue without global visual information, such as only the mouth region, was presented.

Methods

Participants

Another 50 healthy students (12 males and 38 females), who were recruited from an introductory psychology class at Chiba University, participated in the experiment. The mean age of the participants was 19.4 years ($SD = 3.41$). All participants were native speakers of Japanese and reported normal hearing and normal (or corrected) vision. They provided written informed consent in the class and took part in the study voluntarily. After the experiment, they received an oral debriefing.

Stimuli

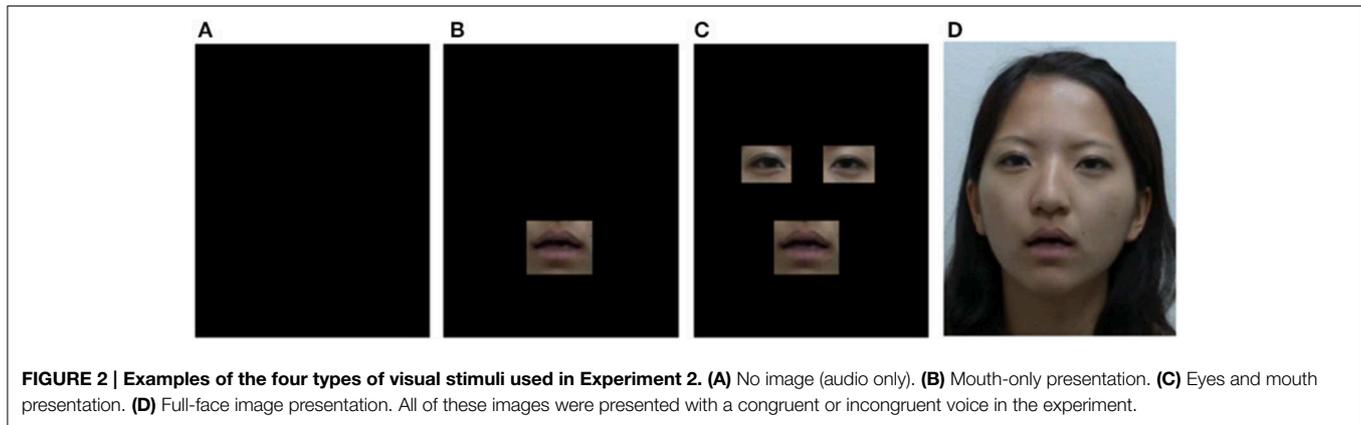
We used the same stimuli as in Experiment 1, comprising six (3 females) Japanese speakers' utterances of three syllables (/pa/, /ta/, or /ka/). There were two audio-visual stimulus conditions, i.e., the audio-visual congruent and audio-visual incongruent. The audio-visual stimuli consisted of a congruent auditory /pa/-visual /pa/, a congruent auditory /ta/-visual /ta/, a congruent auditory /ka/-visual /ka/, and an incongruent auditory /pa/-visual /ka/.

The four types of presentations of visual stimuli—no image (audio-only), mouth-only, eyes and mouth, and full face—(examples of the visual stimuli are shown in **Figure 2**) were created for each condition by using Adobe Premiere Pro CS6 to crop eye regions and the mouth region from visual images. The eye region included the region from the inner corner of the eyes to the outer corner. The mouth region included a range of motion of the upper lip and lower lip. This task consisted of 72 congruent stimuli and 24 incongruent stimuli per block.

Following the experimental tasks, we used the Japanese version of the AQ (Baron-Cohen et al., 2001b; Wakabayashi et al., 2006b) to measure the level of autistic traits in the participants.

Procedure

The experiment was carried out individually, using the same apparatus as in Experiment 1. Participants were seated at a distance of approximately 50 cm from the 19-in CRT monitor, wearing the headphone. They were instructed to report what they heard (/pa/, /ta/, or /ka/) by pressing buttons on a keyboard. In each trial, a fixation point was displayed for 1000 ms at the center of the CRT monitor. After that, either the congruent or the incongruent stimulus was presented, followed by a blank display presented until participants responded. All participants completed the two blocks of the main session after undergoing the 10-trial practice session. The order of trials was randomized



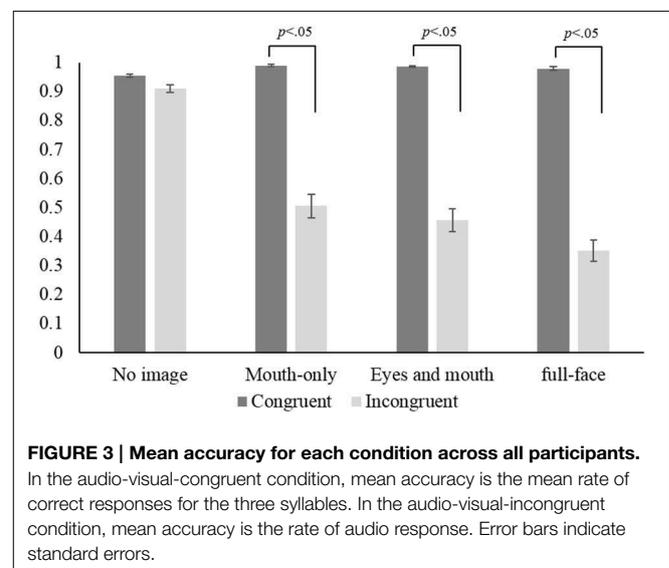
for each block. After the tasks were finished, participants completed the questionnaire.

Data Analysis

Statistical analysis was conducted using R version 2.15.2 for Windows (R Foundation for Statistical Computing, Vienna, Austria). In order to examine the effect of visual presentation type and stimulus condition, rates of correct (audio) responses were analyzed using a Two-Way ANOVA with visual presentation types and stimulus conditions as within-participant factors. As in Experiment 1, the relationship between the AQ scores and task performance was analyzed using Pearson correlation coefficients for each stimulus condition. In addition, group differences between high- and low-AQ groups were analyzed using a mixed ANOVA with visual presentation as a within-participant factor and groups as a between-participants factor for rates of correct responses in the audio-visual-congruent condition and of audio responses in the audio-visual-incongruent condition.

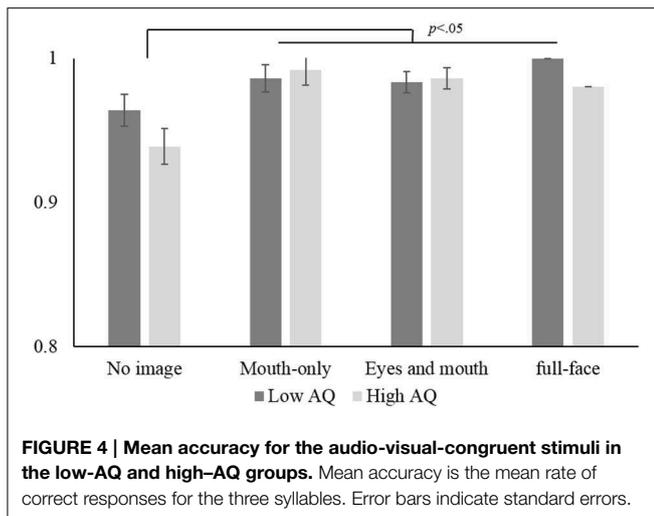
Results

Figure 3 summarizes mean accuracies for the congruent and incongruent stimuli conditions in all types of visual presentation. A Two-Way ANOVA with visual presentation types and stimulus conditions as within-participant factors revealed main effects of stimulus conditions, $F_{(1, 49)} = 201.10, p < 0.01$, partial $\eta^2 = 0.80$, and visual presentation, $F_{(3, 149)} = 110.72, p < 0.01$, partial $\eta^2 = 0.69$, and a significant one-way interaction, $F_{(1, 49)} = 132.10, p < 0.01$, partial $\eta^2 = 0.73$. Multiple comparisons (Holm method) showed that the accuracy for no image, which presented only the auditory stimulus, was lower than for the other types of presentation in the audio-visual-congruent condition ($p < 0.05$). On the other hand, in the audio-visual-incongruent condition, the rate of audio (correct) responses for no image was higher than for the other types of presentation ($p < 0.05$), and the audio response for the full-face presentation was lower than that for either the mouth or mouth and eyes presentation ($p < 0.05$). These results suggested that any type of visual speech cue improved the perception accuracy for audio-visual-congruent stimuli. In addition, as we expected, the influence of a visual speech cue on perceiving a voice was strongest for the presentation of full-face speech with the incongruent voice.



Next, we examined the relationship between the AQ scores and the effect of visual presentation on speech perception. The scores on the AQ ranged from 10 to 43 with a mean of 21.2 ($SD = 6.07$). In order to examine the relationship between task performance and AQ scores, we calculated Pearson correlation coefficients (see Supplementary Material). No significant correlation was observed in the audio-visual-congruent condition. In the audio-visual-incongruent condition, AQ scores were significantly positively correlated with the audio (/pa/) response, $r_{(50)} = 0.31, p < 0.05$, and negatively correlated with the fused (/ta/) response, $r_{(50)} = -0.31, p < 0.05$, but only for the full-face presentation. These correlations replicated the results in Experiment 1, which indicates that individuals with high AQ scores showed less of a visually captured response than individuals with low AQ scores did, although this only occurred in the full-face incongruent speech condition.

Then, we examined group differences between the high-AQ and low-AQ groups in each condition. From among the participants, we picked 10 with scores of 27 or over as the former (2 males and 8 females, mean AQ score = 30.1), and another 10

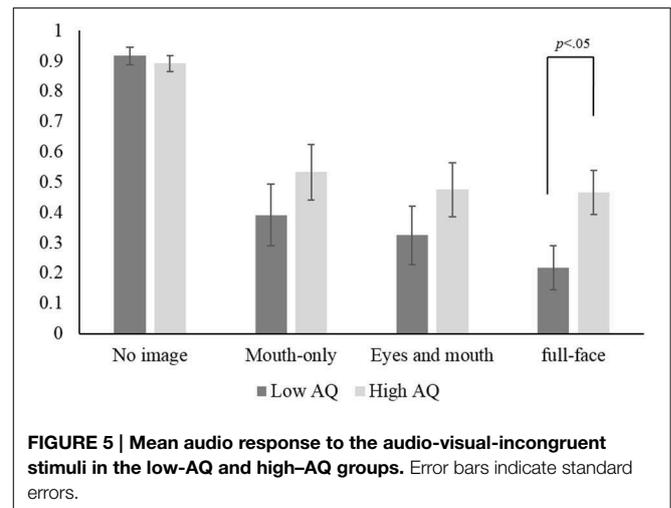


with scores of 16 or under as the latter (5 males and 5 females, mean AQ score = 13.7). A between-groups *t*-test showed a significant difference in the AQ scores, $t_{(18)} = 10.28$, $p < 0.01$, $r = 0.92$. In the congruent condition (see **Figure 4**), a mixed ANOVA revealed that the main effect of visual presentation was significant, $F_{(3, 54)} = 11.42$, $p < 0.01$, partial $\eta^2 = 0.36$, but that the main effect of groups and the one-way interaction were not.

For audio (correct) responses in the incongruent condition (see **Figure 5**), although no significant main effect of groups was found, there was a significant main effect of visual presentation, $F_{(3, 54)} = 63.83$, $p < 0.01$, partial $\eta^2 = 0.75$, and a significant one-way interaction, $F_{(3, 54)} = 3.16$, $p < 0.05$, partial $\eta^2 = 0.15$. This interaction revealed that the simple main effect of groups was significant only in the full-face presentation condition, $F_{(1, 18)} = 5.37$, $p < 0.01$, partial $\eta^2 = 0.23$, so that the individual differences between the high- and low-AQ groups appeared only in the full-face presentation condition. We also found that the effect size of visual presentation was slightly smaller in the high-AQ group, $F_{(3, 54)} = 19.88$, $p < 0.01$, partial $\eta^2 = 0.72$, than in the low-AQ group, $F_{(3, 54)} = 47.11$, $p < 0.01$, partial $\eta^2 = 0.52$. Similar results were found for fused responses in the incongruent condition, i.e., a main effect of visual presentation, $F_{(3, 54)} = 67.49$, $p < 0.01$, partial $\eta^2 = 0.76$, and a significant one-way interaction, $F_{(3, 54)} = 3.14$, $p < 0.05$, partial $\eta^2 = 0.15$. These results indicate that the effect of global facial information was greater in the low-AQ group than in the high-AQ group, although this effect was found in both groups.

Discussion

In Experiment 2, we aimed to investigate the effect of global facial information on audio-visual speech perception, and its relationship with level of autistic traits. With regard to the former purpose, we hypothesized that audio responses would be observed less frequently for the full-face image of a visual speech cue than for the only-mouth image of a visual speech cue in the incongruent condition. As we expected, our results revealed that the rate of audio responses was lower for the full-face image of a visual speech cue than for the other three types of



visual speech cue in the audio-visual-incongruent condition. This indicates that global facial information enhances the influence of a visual cue on perceiving a voice. Unlike previous results (e.g., Rosenblum et al., 2000; Hietanen et al., 2001), the difference in visual influence between the full-face presentation and the only-mouth presentation was robust in our study. Our result directly supports the assumption that the processing of global facial information (extraoral region) might be used for audio-visual speech integration (e.g., Thomas and Jordan, 2004; Eskelund et al., 2015).

With regard to the relationship with autistic traits, our results showed that the individual differences between the high- and low-AQ groups appeared only when a full-face image of a visual speech cue with an incongruent voice was presented. Such a group difference was not found in accuracies for the audio-visual-congruent stimuli. Furthermore, the effect of global facial information in the McGurk effect was small in the high-AQ group. This indicates that the local bias on face processing might play a role in audio-visual speech perception, as well as in the recognition of facial identity (Joseph and Tanaka, 2003; Deruelle et al., 2004) and of facial expression (Kätsyri et al., 2008). These results suggest that the visual influence on perceiving voice was weaker in individuals with high AQ scores than in those with low AQ scores because of the weakness of processing global facial information in the McGurk effect.

General Discussion

Implications for the Dimensional Model of Autism Spectrum Disorder

In two experiments, we examined the link between level of autistic traits and individual differences in audio-visual speech perception. The results demonstrated that level of autistic traits did not correlate with the accuracy for perceiving audio-visual-congruent speech, regardless of the visual speech presentation condition. Moreover, we did not find a correlation between level of autistic traits and the accuracy for perceiving auditory speech,

although individual differences in the audio-visual-incongruent condition, in which the McGurk effect was observed, were related to the degree of autistic traits in the general population. In the audio-visual-incongruent condition, individuals with high AQ scores showed fewer occurrences of the McGurk effect than individuals with low AQ scores did. These results indicate that autistic traits only correlated with the strength of visual influence on perceiving a voice in the audio-visual-incongruent condition.

Our findings have important implications for the dimensional model of ASD, especially for analog studies investigating symptoms of ASD in the general population. With regard to the influence of visual speech cues on perceiving a voice, our results are consistent with those of several previous studies on ASD (De Gelder et al., 1991; Williams et al., 2004; Saalasti et al., 2011; Stevenson et al., 2014). Such atypical processing in individuals with high AQ scores has been reported in the context of perceptual learning (Reed et al., 2011), perspective-taking (Brunye et al., 2012), and lexical effects on speech perception (Stewart and Ota, 2008). As with these studies, our results also support the dimensional model of ASD.

We considered that previous results were mixed due to the heterogeneity of a clinical ASD population in the profile of hyper- and hypo-sensitivity (Woynaroski et al., 2013). To eliminate this factor, we adopted an analog design and used a sample of individuals with TD, who were free from problems with sensory inputs. As we expected, we found significant relationships between level of autistic traits and individual differences in audio-visual speech perception. Some studies showed that it was possible to control the influence of other factors, such as the Big Five personality traits (Wakabayashi et al., 2006a), schizotypal personality (Wakabayashi et al., 2012), and degree of hyper- or hypo-sensitivity (Ujije and Wakabayashi, 2015). These indicate that an analog design might be an effective approach in the investigation of ASD symptoms to control factors other than the degree of autistic traits.

The Role of Global Facial Information in the Occurrence of the McGurk Effect

Our results showed a more robust effect of global facial information in the occurrence of the McGurk effect, as compared to previous studies (e.g., Rosenblum et al., 2000; Hietanen et al., 2001). This indicates that extraoral region of visual speech cues might be used for audio-visual speech integration (Thomas and Jordan, 2004). Our results, however, could not reveal whether global facial information is critical for audio-visual speech perception. One previous study (Jordan and Thomas, 2011) stated that the mouth region of a visual speech cue is important for audio-visual speech perception, which is something we also found in this study, although the extraoral region could also be used. In this study, global facial information did not have a strong effect on audio-visual-congruent speech perception. This means that the accuracy when hearing a voice increased when any type of visual speech cue was presented with a congruent voice, compared to when only a voice was presented. Moreover, in the incongruent condition, the influence of a visual speech sound appeared even when a voice with an incongruent visual speech cue of only a mouth was exhibited. These findings indicate that

information provided by the mouth region is more critical for audio-visual speech perception, than is global facial information.

An issue in this study is that we did not consider the unnaturalness of the stimulation presentation. In this study, a mouth image included a range of motion of the upper lip and lower lip. This image of a visual speech cue was either presented on black background or presented along with other facial parts (eyes or full face). Jordan and Thomas (2011) pointed out that a display that does not obscure all of the face except for the mouth was unnatural. Therefore, rather than global face processing, this unnaturalness might have been what influenced audio-visual speech perception. Another issue is that we used only one combination of visual and audio syllables in the incongruent condition. Previous studies have shown that the effect of global face processing varied with the stimulus, such as a different talker or a different combination of syllables (Jordan and Bevan, 1997; Rosenblum et al., 2000). Nevertheless, the number of talkers in our stimuli (six talkers) was relatively larger than that used in previous studies (Jordan and Bevan, 1997; Rosenblum et al., 2000), as was our sample size.

The Relationship between Level of Autistic Traits and Local Bias in the McGurk Effect

With regard to the local bias exhibited by individuals with ASD, our results suggested a link between level of autistic traits and a weak ability to process global facial information in McGurk stimuli. In our results, the effect of global facial information in the McGurk stimuli was found to be smaller in individuals with low AQ scores than in individuals with high AQ scores, who show less likelihood of the McGurk effect occurring. This could be interpreted as indicating that individuals with high AQ scores show a local bias toward a visual speech cue and that their weak ability to process global facial information leads to individual differences in the McGurk effect. On the other hand, it is possible that other factors might have influenced our results, such as the atypical processing of global motion (Koldewyn et al., 2010), of visual attention (Zhao et al., 2013), or of gaze behavior (Klin et al., 2002).

Previous results provide us with some advantage to understand the influence of gaze behavior in our study. It has been shown that individuals with ASD exhibit atypical gaze behavior toward faces when they observe face stimuli (see, for a review, Senju and Johnson, 2009). Klin et al. (2002) indicated that individuals with ASD tend to fixate more on the mouth region when a dynamic face is presented, while individuals without ASD tend to fixate more on the region of the eyes. However, Saalasti et al. (2012) reported that no differences in gaze behavior between adults with ASD and controls that could have accounted for the individual differences in the McGurk effect. Moreover, Paré et al. (2003) showed that gaze fixations within the talker's face, which meant that gaze was fixed on the talker's mouth or on the talker's eyes, did not influence the likelihood of the McGurk effect occurring in adults with TD. Thus, it could be considered that, even if gaze behavior during trials differed between the high-AQ and low-AQ groups in Experiment 2, such

differences would not have substantially influenced the results of this study.

As another limitation in this study, it was unclear whether the individual differences in lip-reading were related to individual differences in the McGurk effect, because we did not use visual-only stimuli in this experiment. Some studies have reported that individuals with ASD experience a deficit in perceiving audio-visual speech because of their poor ability to lip-read (Williams et al., 2004; Woynaroski et al., 2013). Therefore, if individuals with high AQ scores have difficulties in lip-reading, individual differences in the McGurk effect might be caused by poor lip-reading ability, rather than by a local bias toward a visual speech cue. Nevertheless, the results of Experiment 2 showed a significant main effect of visual presentation in the congruent condition for both the high-AQ and low-AQ groups. In other words, when any type of congruent visual speech cue was exhibited, improved accuracy for perceiving a voice was found regardless of level of AQ. If the high-AQ group in this study had difficulties in lip-reading, such improvement would not have been found in that group. In order to clarify the role of a local bias toward a visual speech cue during audio-visual speech perception, these factors should be investigated directly in further studies.

References

- American Psychiatric Association. (1994). *Diagnostic and Statistical Manual of Mental Disorders, 4th Edn.* Washington, DC: American Psychiatric Association.
- American Psychiatric Association. (2013). *Diagnostic and Statistical Manual of Mental Disorders, 5th Edn.* Washington, DC: American Psychiatric Association.
- Baron-Cohen, S. (1995). *Mindblindness: An Essay on Autism and Theory of Mind.* Boston, MA: MIT Press/Bradford Books.
- Baron-Cohen, S., Wheelwright, S., Hill, J., Raste, Y., and Plumb, I. (2001a). The 'Reading the Mind in the eyes' test revised version: a study with normal adults, and adults with Asperger Syndrome or High-Functioning Autism. *J. Child Psychol. Psychiatry* 42, 241–252. doi: 10.1111/1469-7610.00715
- Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J., and Clubley, E. (2001b). The Autism-Spectrum Quotient (AQ): evidence from Asperger syndrome/high-functioning autism, males and females, scientists and mathematicians. *J. Autism Dev. Disord.* 31, 5–17. doi: 10.1023/A:1005653411471
- Brosnan, M. J., Gwilliam, L. R., and Walker, I. (2012). Brief report: the relationship between visual acuity, the embedded figures test and systemizing in autism spectrum disorders. *J. Autism Dev. Disord.* 42, 2491–2497. doi: 10.1007/s10803-012-1505-0
- Bruce, V., and Young, A. (1986). Understanding face recognition. *Br. J. Psychol.* 77, 305–327. doi: 10.1111/j.2044-8295.1986.tb02199.x
- Brune, T. T., Ditman, T., Giles, G. E., Mahoney, C. R., Kessler, K., and Taylor, H. A. (2012). Gender and autistic personality traits predict perspective-taking ability in typical adults. *J. Pers. Individ. Dif.* 52, 84–88. doi: 10.1016/j.paid.2011.09.004
- Calvert, G. A., Campbell, R., and Brammer, M. J. (2000). Evidence from functional magnetic resonance imaging of crossmodal binding in the human heteromodal cortex. *Curr. Biol.* 10, 649–657. doi: 10.1016/S0960-9822(00)00513-3
- Crane, L., Goddard, L., and Pring, L. (2009). Sensory processing in adults with autism spectrum disorders. *Autism* 13, 215–228. doi: 10.1177/1362361309103794
- De Gelder, B., Vroomen, J., and van der Heide, L. (1991). Face recognition and lip-reading in autism. *Eur. J. Cogn. Psychol.* 3, 69–86. doi: 10.1080/09541449108406220

Conclusion

In conclusion, level of the autistic traits in the general population was found to correlate negatively with visually influenced percepts with the McGurk stimuli. This is the first report of such a correlation. Moreover, individuals with high levels of autistic traits showed a weak ability to process global facial information during the McGurk stimuli.

Acknowledgments

We would like to thank Dr. A. Tanaka for advising us about the process of creating an audio-visual speech task. We would also like to thank all the students for voluntarily participating in our experiments. This study was supported by Grant-in-Aid from the Japan Society for the Promotion of Science Fellows (Grant No. 26-8144).

Supplementary Material

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fpsyg.2015.00891>

- Deruelle, C., Rondan, C., Gepner, B., and Tardif, C. (2004). Spatial frequency and face processing in children with autism and Asperger syndrome. *J. Autism Dev. Disord.* 34, 199–210. doi: 10.1023/B:JADD.0000022610.09668.4c
- Dunn, W., and Westman, K. (1997). The sensory profile: the performance of a national sample of children without disabilities. *Am. J. Occup. Ther.* 51, 25–34. doi: 10.5014/ajot.51.1.25
- Eskehlund, K., MacDonald, E. N., and Andersen, T. S. (2015). Face configuration affects speech perception: evidence from a McGurk mismatch negativity study. *Neuropsychologia* 66, 48–54. doi: 10.1016/j.neuropsychologia.2014.10.021
- Foss-Feig, J. H., Heacock, J. L., and Cascio, C. J. (2012). Tactile responsiveness patterns and their association with core feature in autism spectrum disorders. *Res. Autism Spectr. Disord.* 6, 337–344. doi: 10.1016/j.rasd.2011.06.007
- Frith, U. (1991). *Autism and Asperger's Syndrome.* Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511526770
- Haesen, B., Boets, B., and Wagemans, J. (2011). A review of behavioral and electrophysiological studies on auditory processing and speech perception in autism spectrum disorders. *Res. Autism Spectr. Disord.* 5, 701–714. doi: 10.1016/j.rasd.2010.11.006
- Happé, F., and Frith, U. (2006). The weak coherence account: detail-focused cognitive style in autism spectrum disorders. *J. Autism Dev. Disord.* 36, 5–25. doi: 10.1007/s10803-005-0039-0
- Hietanen, J. K., Manninen, P., Sams, M., and Surakka, V. (2001). Does audiovisual speech perception use information about facial configuration? *Eur. J. Cogn. Psychol.* 13, 395–407. doi: 10.1080/09541440126006
- Hoekstra, R. A., Bartels, M., Cath, D. C., and Boomsma, D. I. (2008). Factor structure, reliability and criterion validity of the Autism-Spectrum Quotient (AQ): a study in Dutch population and patient groups. *J. Autism Dev. Disord.* 38, 1555–1566. doi: 10.1007/s10803-008-0538-x
- Iarocci, G., Rombough, A., Yager, J., Weeks, D., and Chua, R. (2010). Visual influences on speech perception in children with autism. *Autism* 14, 305–320. doi: 10.1177/1362361309353615
- Jordan, T. R., and Bevan, K. (1997). Seeing and hearing rotated faces: influences of official orientation on visual and audiovisual speech recognition. *J. Exp. Psychol. Hum. Percept. Perform.* 23, 388–403. doi: 10.1037/0096-1523.23.2.388
- Jordan, T. R., and Thomas, S. M. (2011). When half a face is as good as a whole: effects of simple substantial occlusion on visual and audiovisual speech

- perception. *Atten. Percept. Psychophys.* 73, 2270–2285. doi: 10.3758/s13414-011-0152-4
- Joseph, R. M., and Tanaka, J. (2003). Holistic and part-based face recognition in children with autism. *J. Child Psychol. Psychiatry* 44, 529–542. doi: 10.1111/1469-7610.00142
- Kätsyri, J., Saalasti, S., Tiippana, K., von Wendt, L., and Sams, M. (2008). Impaired recognition of facial emotions from low-spatial frequencies in Asperger syndrome. *Neuropsychologia* 46, 1888–1897. doi: 10.1016/j.neuropsychologia.2008.01.005
- Keane, B. P., Rosenthal, O., Chun, N. H., and Shams, L. (2010). Audiovisual integration in high functioning adults with autism. *Res. Autism Spectr. Disord.* 4, 276–289. doi: 10.1016/j.rasd.2009.09.015
- Klin, A., Jones, W., Schultz, R., Volkmar, R., and Cohen, D. (2002). Visual fixation patterns during viewing of naturalistic social situations as predictors of social competence in individuals with autism. *Arch. Gen. Psychiatry* 59, 809–816. doi: 10.1001/archpsyc.59.9.809
- Koldewyn, K., Whitney, D., and Rivera, S. M. (2010). The psychophysics of visual motion and global form processing in autism. *Brain* 133, 599–610. doi: 10.1093/brain/awp272
- Lau, W. Y. P., Kelly, A. B., and Peterson, C. C. (2013). Further evidence on the factorial structure of the autism spectrum quotient (AQ) for adults with and without a clinical diagnosis of autism. *J. Autism Dev. Disord.* 43, 2807–2815. doi: 10.1007/s10803-013-1827-6
- Massaro, D. W., and Bosseler, A. (2006). Read my lips: the importance of the face in a computer-animated tutor for vocabulary learning by children with autism. *Autism* 10, 495–510. doi: 10.1177/1362361306066599
- McGurk, H., and MacDonald, J. W. (1976). Hearing lips and seeing voices. *Nature* 264, 746–748. doi: 10.1038/264746a0
- Nath, A. R., and Beauchamp, M. S. (2011). Dynamic changes in superior temporal sulcus connectivity during perception of noisy audiovisual speech. *J. Neurosci.* 31, 1704–1714. doi: 10.1523/JNEUROSCI.4853-10.2011
- Nath, A. R., and Beauchamp, M. S. (2012). A neural basis for interindividual differences in the McGurk effect, a multisensory speech illusion. *Neuroimage* 59, 781–787. doi: 10.1016/j.neuroimage.2011.07.024
- O'Connor, K. (2012). Auditory processing in autism spectrum disorder: a review. *Neurosci. Biobehav. Rev.* 36, 836–854. doi: 10.1016/j.neubiorev.2011.11.008
- Paré, M., Richler, R., ten Hove, M., and Munhall, K. G. (2003). Gaze behavior in audiovisual speech perception: the influence of ocular fixations on the McGurk effect. *Percept. Psychophys.* 65, 553–567. doi: 10.3758/BF03194582
- Redcay, E. (2008). The superior temporal sulcus performs a common function for social and speech perception: implications for the emergence of autism. *Neurosci. Biobehav. Rev.* 32, 123–142. doi: 10.1016/j.neubiorev.2007.06.004
- Reed, P., Lowe, C., and Everett, R. (2011). Perceptual learning and perceptual search are altered in male university students with higher Autism Quotient scores. *Pers. Individ. Dif.* 51, 732–736. doi: 10.1016/j.paid.2011.06.016
- Robertson, A. E., and Simmons, D. R. (2012). The relationship between sensory sensitivity and autistic traits in the general population. *J. Autism Dev. Disord.* 43, 775–784. doi: 10.1007/s10803-012-1608-7
- Rosenblum, L. D., Yakel, D. A., Baseer, N., Panchal, A., Nodarse, B. C., and Niehus, R. P. (2002). Visual speech information for face recognition. *Percept. Psychophys.* 64, 220–229. doi: 10.3758/BF03195788
- Rosenblum, L. D., Yakel, D. A., and Green, K. G. (2000). Face and mouth inversion affects on visual and audiovisual speech perception. *J. Exp. Psychol. Hum. Percept. Perform.* 26, 806–819. doi: 10.1037/0096-1523.26.2.806
- Saalasti, S., Kätsyri, J., Tiippana, K., Laine-Hernandez, M., von Wendt, L., and Sams, M. (2012). Audiovisual speech perception and eye gaze behavior of adults with Asperger syndrome. *J. Autism Dev. Disord.* 42, 1606–1615. doi: 10.1007/s10803-011-1400-0
- Saalasti, S., Tiippana, K., Kätsyri, J., and Sams, M. (2011). The effect of visual spatial attention on audiovisual speech perception in adults with Asperger syndrome. *Exp. Brain Res.* 213, 283–290. doi: 10.1007/s00221-011-2751-7
- Sekiya, K. (1994). Difference in auditory-visual speech perception between Japanese and Americans: McGurk effect as a function of incompatibility. *J. Acoust. Soc. Jpn.* 15, 143–158. doi: 10.1250/ast.15.143
- Senju, A., and Johnson, M. H. (2009). Atypical eye contact in autism: models, mechanisms and development. *Neurosci. Biobehav. Rev.* 33, 1204–1214. doi: 10.1016/j.neubiorev.2009.06.001
- Simmons, D. R., Robertson, A. E., McKay, L. S., Toal, E., McAleer, P., and Pollick, F. E. (2009). Vision in autism spectrum disorders. *Vis. Res.* 49, 2705–2739. doi: 10.1016/j.visres.2009.08.005
- Smith, E. G., and Bennetto, L. (2007). Audiovisual speech integration and lipreading in autism. *J. Child Psychol. Psychiatry* 48, 813–821. doi: 10.1111/j.1469-7610.2007.01766.x
- Stevenson, R. A., Siemann, J. K., Woynarowski, T. G., Schneider, B. C., Eberly, H. E., Camarata, S. M., et al. (2014). Brief report: arrested development of audiovisual speech perception in autism spectrum disorders. *J. Autism Dev. Disord.* 44, 1470–1477. doi: 10.1007/s10803-013-1992-7
- Stewart, M. E., and Ota, M. (2008). Lexical effects on speech perception in individuals with autistic traits. *Cognition* 109, 157–162. doi: 10.1016/j.cognition.2008.07.010
- Suda, M., Takei, Y., Aoyama, Y., Narita, K., Sakurai, N., Fukuda, M., et al. (2011). Autistic traits and brain activation during face-to-face conversations in typically developed adults. *PLoS ONE* 6:e20021. doi: 10.1371/journal.pone.0020021
- Sumbly, W. H., and Pollack, I. (1954). Visual contribution to speech intelligibility in noise. *J. Acoust. Soc. Am.* 26, 212–215. doi: 10.1121/1.1907309
- Taylor, N., Isaac, C., and Milne, E. (2010). A comparison of the development of audiovisual integration in children with autism spectrum disorders and typically developing children. *J. Autism Dev. Disord.* 40, 1403–1411. doi: 10.1007/s10803-010-1000-4
- Thomas, S. M., and Jordan, T. R. (2004). Contributions of oral and extraoral facial movement to visual and audiovisual speech perception. *J. Exp. Psychol. Hum. Percept. Perform.* 30, 873–888. doi: 10.1037/0096-1523.30.5.873
- Tomchek, S. D., and Dunn, W. (2007). Sensory processing in children with and without autism: a comparative study using the short sensory profile. *Am. J. Occup. Ther.* 61, 190–200. doi: 10.5014/ajot.61.2.190
- Ujjie, Y., and Wakabayashi, A. (2015). Psychometric properties and overlap of the GSQ and AQ among Japanese university students. *Int. J. Psychol. Stud.* 7, 195–205. doi: 10.5539/ijps.v7n2p195
- Wakabayashi, A., Baron-Cohen, S., and Ashwin, C. (2012). Do the traits of autism-spectrum overlap with those of schizophrenia or obsessive-compulsive disorder in the general population? *Res. Autism Spectr. Disord.* 6, 717–725. doi: 10.1016/j.rasd.2011.09.008
- Wakabayashi, A., Baron-Cohen, S., and Wheelwright, S. (2006a). Are autistic traits an independent personality dimension? A study of the Autism-Spectrum Quotient (AQ) and the NEO-PI-R. *J. Pers. Individ. Dif.* 41, 873–883. doi: 10.1016/j.paid.2006.04.003
- Wakabayashi, A., Baron-Cohen, S., Wheelwright, S., and Tojo, Y. (2006b). The Autism-Spectrum Quotient (AQ) in Japan: a crosscultural comparison. *J. Autism Dev. Disord.* 36, 263–270. doi: 10.1007/s10803-005-0061-2
- Williams, J. H. G., Massaro, D. W., Peel, N. J., Bosseler, A., and Suddendorf, T. (2004). Visual-auditory integration during speech imitation in autism. *Res. Dev. Disabil.* 25, 559–575. doi: 10.1016/j.ridd.2004.01.008
- Woynarowski, T. G., Kwakye, L. D., Foss-Feig, J. H., Stevenson, R. A., Stone, W. L., and Wallace, M. T. (2013). Multisensory speech perception in children with autism spectrum disorders. *J. Autism Dev. Disord.* 43, 2891–2902. doi: 10.1007/s10803-013-1836-5
- Zhao, S., Uono, S., Yoshimura, S., Kubota, Y., and Toichi, M. (2013). Can gaze-cueing be helpful for detecting sound in autism spectrum disorder? *Res. Autism Spectr. Disord.* 7, 1250–1256. doi: 10.1016/j.rasd.2013.07.001
- Zilbovicius, M., Meresse, I., Chabane, N., Brunelle, F., Samson, Y., and Boddaert, N. (2006). Autism, the superior temporal sulcus and social perception. *Trends Neurosci.* 29, 359–366. doi: 10.1016/j.tins.2006.06.004

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2015 Ujjie, Asai and Wakabayashi. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.