

# Weaker McGurk Effect for Rubin's Vase-Type Speech in People With High Autistic Traits

Yuta Ujiie<sup>1,2,3,\*\*</sup> and Kohske Takahashi<sup>4,\*\*\*</sup>

<sup>1</sup> Graduate School of Psychology, Chukyo University, 101-2 Yagoto Honmachi, Showa-ku, Nagoya-shi, Aichi, 466-8666, Japan

<sup>2</sup> Japan Society for the Promotion of Science, Kojimachi Business Center Building, 5-3-1 Kojimachi, Chiyoda-ku, Tokyo 102-0083, Japan

<sup>3</sup> Research and Development Initiative, Chuo University, 1-13-27, Kasuga, Bunkyo-ku, Tokyo, 112-8551, Japan

<sup>4</sup> School of Psychology, Chukyo University, 101-2 Yagoto Honmachi, Showa-ku, Nagoya-shi, Aichi, 466-8666, Japan

Received 18 August 2020; accepted 5 April 2021

---

## Abstract

While visual information from facial speech modulates auditory speech perception, it is less influential on audiovisual speech perception among autistic individuals than among typically developed individuals. In this study, we investigated the relationship between autistic traits (Autism-Spectrum Quotient; AQ) and the influence of visual speech on the recognition of Rubin's vase-type speech stimuli with degraded facial speech information. Participants were 31 university students (13 males and 18 females; mean age: 19.2, SD: 1.13 years) who reported normal (or corrected-to-normal) hearing and vision. All participants completed three speech recognition tasks (visual, auditory, and audiovisual stimuli) and the AQ–Japanese version. The results showed that accuracies of speech recognition for visual (i.e., lip-reading) and auditory stimuli were not significantly related to participants' AQ. In contrast, audiovisual speech perception was less susceptible to facial speech perception among individuals with high rather than low autistic traits. The weaker influence of visual information on audiovisual speech perception in autism spectrum disorder (ASD) was robust regardless of the clarity of the visual information, suggesting a difficulty in the process of audiovisual integration rather than in the visual processing of facial speech.

## Keywords

autism-spectrum quotient, audiovisual speech, McGurk effect, Rubin's vase

---

\* To whom correspondence should be addressed. E-mail: yuta.ujiie.160330@gmail.com

\*\* ORCID: 0000-0003-0637-7880

\*\*\* ORCID: 0000-0001-6076-4828

## 1. Introduction

### 1.1. McGurk Effect

Audiovisual speech perception is an adaptive function in face-to-face communication, which helps individuals to recognize another person's speech, especially in environments where it is difficult to clearly hear the voice (e.g., Liberman and Mattingly, 1985; Sumbly and Pollack, 1954). The interaction of voice and facial speech can also be seen in a well-known illusion, the McGurk effect (McGurk and MacDonald, 1976), in which a facial speech cue (/ka/) dubbed with an incongruent voice (/pa/) is presented, and most observers will perceive an intermediate phoneme (/ta/) even if they are instructed to ignore the facial speech cue. In this process, contradictory phonetic and visual information is loaded simultaneously and then processed interactively through an integration process, which results in a fused percept (e.g., Beauchamp *et al.*, 2004; Calvert *et al.*, 2000). The extent of susceptibility to the McGurk effect is considered to reflect the degree of a perceiver's reliance on facial speech during audiovisual speech processing (e.g., Rosenblum, 2008; Tiippana, 2014).

The processing of facial speech is essential for the occurrence of the McGurk effect (e.g., Rosenblum and Saldaña, 1996; Rosenblum *et al.*, 2000, 2002). Facial speech comprises both kinematic and structural facial information. Kinematic information refers to the facial movements when articulating syllables. The essential role of kinematic information in the McGurk effect was demonstrated by using a point-light display of facial speech from which syllable or speech information could be visually restored (e.g., Jordan *et al.*, 2000; Rosenblum and Saldaña, 1996; Rosenblum *et al.*, 2002). In contrast, structural facial information refers to compositions of facial parts in the facial contour (e.g., Eskelund *et al.*, 2015; Hietanen *et al.*, 2001). Some studies have shown that the McGurk effect decreases if structural facial information is distorted: a reduced McGurk effect was demonstrated with inverted face stimuli (Jordan and Bevan, 1997; Ujiie *et al.*, 2018) and Thatcher-illusion face stimuli (Eskelund *et al.*, 2015; Rosenblum *et al.*, 2000; Ujiie *et al.*, 2018).

### 1.2. A Weaker McGurk Effect in Autism Spectrum Disorder

Susceptibility to the McGurk effect is decreased in autistic individuals compared to typically developed (TD) individuals (e.g., de Gelder *et al.*, 1991; Iarocci *et al.*, 2010; Saalasti *et al.*, 2012; van Laarhoven *et al.*, 2019; Williams *et al.*, 2004). Autistic individuals are likely to be diagnosed with autism spectrum disorder (ASD), which is a neurodevelopmental disorder characterized by difficulties with social interaction and communication, patterns of repetitive behavior, narrow interests, and difficulties with sensory processing [Diagnostic and Statistical Manual of Mental Disorders-5 (DSM-5); American

Psychiatric Association, 2013]. Previous studies suggest that during audio-visual speech perception, autistic individuals show a weaker McGurk effect as they rely less, or may not rely at all, on facial speech (e.g., de Gelder *et al.*, 1991; Iarocci *et al.*, 2010; Saalasti *et al.*, 2012; Williams *et al.*, 2004). A meta-analysis of nine clinical studies of the McGurk effect supported the findings of the aforementioned studies that individuals diagnosed with ASD exhibit a weaker McGurk effect than do TD control groups (Zhang *et al.*, 2019) whereas other studies reported mixed results, in which autistic individuals demonstrated a McGurk effect at the same level as their counterparts (typically developing individuals; e.g., Keane *et al.*, 2010; Woynaroski *et al.*, 2013).

A weaker McGurk effect has also been observed in individuals with high autistic traits who were not diagnosed with ASD (e.g., Ujiie and Wakabayashi, 2019; Ujiie *et al.*, 2015, 2018; van Laarhoven *et al.*, 2019). The dimensional model of ASD assumes that the degree of autistic-like traits is distributed on a continuum across the general and clinical populations (e.g., Baron-Cohen, 1995, 2002; Frith, 1991), which also enables investigation of perceptual and behavioral tendencies in ASD using a sample of typically developing persons (i.e., analogous study). Previous analogous studies of ASD that measured autistic traits using the Autism-Spectrum Quotient (AQ; Baron-Cohen *et al.*, 2001a) found that people with high AQs but no diagnosis of ASD showed similar tendencies in perceptual and cognitive tasks to people who were diagnosed with ASD (Baron-Cohen *et al.*, 2001b; Reed *et al.*, 2011; Stewart and Ota, 2008). Consistent with this, Ujiie *et al.* (2015) tested the relationship between the McGurk effect and autistic traits in a typically developing sample; higher levels of autistic traits were associated with fewer fused responses and more auditory responses to McGurk stimuli. Similar results were observed in subsequent studies (e.g., Ujiie and Wakabayashi, 2019; Ujiie *et al.*, 2018; van Laarhoven *et al.*, 2019). These results suggest that greater levels of autistic traits are associated with a weaker McGurk effect, even in typically developing adults (Bebko *et al.*, 2014).

A weaker McGurk effect in autistic individuals may be associated with the processing of visual aspects of facial information. Although the McGurk effect is a phenomenon that derives from multisensory integration processes, processing of both kinematic and structural facial information contributes to the occurrence of the McGurk effect (e.g., Rosenblum and Saldaña, 1996; Rosenblum *et al.*, 2000, 2002). In addition, previous studies of ASD have revealed atypical processing of structural facial information (Deruelle *et al.*, 2004; Kättsyri *et al.*, 2008; Kikuchi *et al.*, 2013) and point-light movement perception (Koldewyn *et al.*, 2011; O'Brien *et al.*, 2014), although mixed results also exist (e.g., Rouse *et al.*, 2004; Teunisse and de Gelder, 2001; see Tanaka and Sung, 2016 for a review). Our prior studies examined whether a weak McGurk

effect in ASD is caused by the atypical processing of structural facial information, by manipulating the types of visual stimuli presented (parts of the face or the full face in Ujiie *et al.*, 2015; misoriented facial parts in Ujiie *et al.*, 2018). The results demonstrated that the McGurk effect was weaker in individuals with high AQ scores than in those with low AQ scores, even when distorted structural facial information was used in McGurk stimuli (Ujiie *et al.*, 2015, 2018). This suggests that the difficulties with processing structural facial information were the main cause of the weaker McGurk effect in autistic individuals.

### 1.3. Purpose of the Present Study

In the present study, we hypothesized that a weaker McGurk effect in autistic individuals is associated with the processing of kinematic information. To test this hypothesis, we used Rubin vase-type audiovisual speech stimuli (Munhall *et al.*, 2009) in an analogous study. The Rubin vase image containing bistable figures (profiles of faces or a vase) can be consciously perceived as face-like features and would preserve kinematic information, even though structural facial information such as facial components and textures are degraded (e.g., Hasson *et al.*, 2001; Hesselmann *et al.*, 2008; Peterson *et al.*, 1991). Munhall *et al.* (2009) developed novel stimuli for the McGurk effect by utilizing Rubin's vase images. In their results, Rubin's vase-type visual speech, which contains the moving profiles of two faces (or a vase), produced the McGurk effect when visual speech was presented with an incongruent voice. In relation to Rubin's vase and ASD, a previous study showed that sensitivity to face-like features in the Rubin's vase illusion is comparable between autistic individuals and controls (Akechi *et al.*, 2014). This suggests that the ability to detect facial speech from Rubin's vase-type visual speech would not differ with levels of autistic traits.

In the experiment, participants engaged in the visual, auditory, and audiovisual speech recognition task. Following our hypothesis and research context discussed above, we predicted that people with high AQ scores would demonstrate weak influence of visual speech (i.e., weak McGurk effect) for Rubin's vase-type McGurk stimuli, whereas the performances of the auditory and visual speech recognition would be independent of the participants' AQ scores.

## 2. Methods

### 2.1. Participants

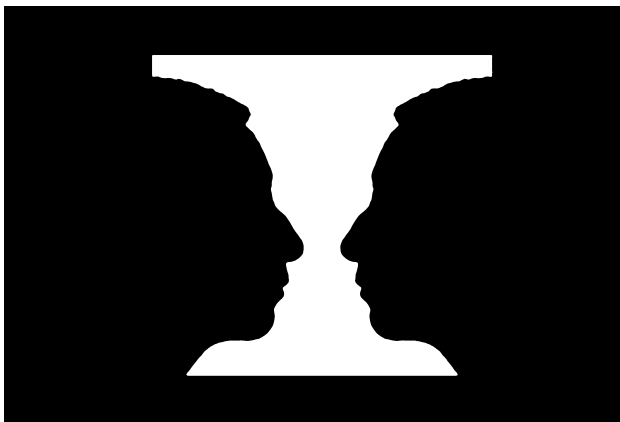
Thirty-one university students (13 males and 18 females; mean age: 19.2, SD: 1.13 years), none of whom had any diagnosis of developmental disorders, participated in this experiment. They were recruited from one university through

a cloud-based participant management system (Sona Systems). We considered that our sample ( $n = 31$ ) had a sufficient size to show a normal distribution of AQ. A normal distribution of AQ has been repeatedly found in the Japanese population (e.g., Wakabayashi *et al.*, 2006; Ujiie and Wakabayashi, 2015). Participants were naïve to the purpose of the experiment. All the participants were native Japanese speakers and reported normal (or corrected) hearing and vision. Informed consent was obtained from all participants. This study was approved by the local ethical committee of Chukyo University in accordance with the standards of the Declaration of Helsinki.

## 2.2. Materials

### 2.2.1. Audiovisual Speech Recognition Task

We created the stimuli combining simultaneous audio and video recordings of two Japanese speakers' utterances (one female) of three syllables ('apa', 'ata' and 'aka'). The voices (digitized at 48 kHz, with a 16-bit quantization resolution) were recorded using a dynamic microphone (MD42, Sennheiser, Wedemark, Germany). The visual stimuli ( $720 \times 576$  pixels, 29.97 frames/s) were recordings of the right or left side of the speakers' faces using a digital video camera (GZ-EX370, JVC Kenwood, Yokohama, Japan). Rubin's vase-type movies were created by combining the speaker's faces on the original and horizontally flipped side for each utterance, using Adobe Premiere Pro CS6 (Adobe Inc., San Jose, CA, USA; Fig. 1). In the movies, both face silhouettes were facing each other and moving while simultaneously pronouncing the same syllable (e.g., 'aka'). Then, we created the audiovisual stimuli with synchronized audio and visual stimuli for each syllable. The audiovisual stimuli included six congruent stimuli (three syllables, from two speakers) and two incongruent (McGurk-type) stimuli (a combination of audio 'apa' and visual



**Figure 1.** An example of a visual stimulus.

‘aka’ for each pair of speakers). Pink noise was added (the signal-to-noise ratio was +10 dB) to enhance the impact of visual speech on voice perception. Additionally, we created visual stimuli without any voices, and auditory stimuli without any faces. The mean duration of the audiovisual stimuli was 2.0 s. The final audiovisual speech recognition task comprised four stimulus conditions: auditory condition (six auditory stimuli), visual condition (six visual stimuli), congruent condition (six congruent stimuli), and the McGurk condition (two incongruent stimuli).

### 2.2.2. *Autism-Spectrum Quotient (AQ–Japanese Version)*

We used the AQ–Japanese version (Wakabayashi *et al.*, 2006) to measure autistic traits in adults with a normal IQ. The AQ was developed by Baron-Cohen *et al.* (2001a) based on the dimensional ASD model (Baron-Cohen, 1995; Frith, 1991). The psychometric properties of the AQ were reported for the original sample, including score distribution, test reliability, and validity for screening between clinical and control groups (in the UK; Baron-Cohen *et al.*, 2001b) and replicated in different samples (e.g., in the Netherlands: Hoekstra *et al.*, 2008; in Australia: Lau *et al.*, 2013; and in Japan: Wakabayashi *et al.*, 2006). The AQ consists of 50 items assessing five ASD symptom domains: social skills, attention switching, attention to detail, communication, and imagination. The participants rate each item by choosing one of the following responses: ‘definitely agree’, ‘slightly agree’, ‘slightly disagree’, and ‘definitely disagree’. Each item receives a score of 0 or 1, according to the scoring method used in previous studies (Baron-Cohen *et al.*, 2001a; Wakabayashi *et al.*, 2006), and the total score ranges from 0 to 50. Previous studies suggested a cutoff score to discriminate clinical and nonclinical populations of  $AQ > 27$  (Woodbury-Smith *et al.*, 2005); persons who meet this criterion are referred to as exhibiting a broader autism phenotype (e.g., Austin, 2005; Hasegawa *et al.*, 2015).

### 2.3. *Procedure*

Participants were seated approximately 50 cm from a 23.5-inch (59.7 cm) cathode-ray tube monitor (FG2421, EIZO, Ishikawa, Japan). The visual angle subtended by the Rubin’s vase stimulus was  $10.67 \times 10.07$  degrees. The voice and pink noise were presented via headphones (MDR-Z500, Sony, Tokyo, Japan) at approximately 65 dB SPL, adjusted using a mixing console (MW8CX, Yamaha, Hamamatsu, Japan). The experiment was conducted using Inquisit version 4.0.6 (Millisecond Software, Seattle, WA, USA). The experiment comprised three blocks: (1) an auditory block consisting of six auditory stimuli, (2) an audiovisual block consisting of six congruent and two McGurk stimuli, and (3) a visual block consisting of six visual stimuli. Each stimulus was presented five times per block. In each trial, a fixation point was

displayed for 1,000 ms at the center of the monitor, followed by the stimulus. In the auditory and audiovisual blocks, the experimenter asked the participants “what did you hear?”, while in the visual block, the experimenter asked “what did you think he was saying?” In the instruction phase, the participants were told that visual stimuli were facial speech articulating one of three syllables. As per the instructions, the participants responded appropriately by pressing a key corresponding to ‘apa’, ‘ata’, or ‘aka’. The order of blocks was counter-balanced across participants. The order of the trials was randomized in each block. All participants completed the main session of three blocks after a practice session of 10 trials comprising the audiovisual congruent and McGurk stimuli. After they finished the experimental task, they completed the AQ–Japanese version.

#### 2.4. Analysis

The statistical analysis was conducted using R version 3.6.2 for Windows (R Foundation for Statistical Computing, Vienna, Austria). We calculated mean correct responses for the auditory, visual, and audiovisual congruent stimuli. For the McGurk-type stimuli, there were three possible responses: audio (‘apa’ response), fused (‘ata’ response), and visual (‘aka’ response). We considered the mean audio response rate in the McGurk condition (i.e., the number of responses correctly judging the sound as ‘apa’) as an indicator of the weakness of visual influence on audio perception (e.g., MacDonald, 2018; Ujiie and Wakabayashi, 2019; Ujiie *et al.*, 2015, 2018).

The relationship between AQ and task performance was analyzed using Pearson’s correlation coefficient for each stimulus condition. Since previous studies reported that the five-factor model of AQ is not reliable (e.g., Austin, 2005; Hoekstra *et al.*, 2008), we used the total AQ score and not each subscore of the analysis. In addition, mean correct responses for the audiovisual congruent stimuli and audio responses for McGurk-type stimuli were analyzed using a multiple regression analysis with stimulus type (congruent and McGurk-type), AQ, and their interaction as predictor variables. Although most previous studies focused on the ratio of fused responses (i.e., ‘ata’) or visually influenced responses (i.e., ‘ata’ and ‘aka’; reviewed in Zhang *et al.*, 2019), we used auditory responses to the McGurk stimuli in the multiple regression analysis to test whether autistic traits differed between congruent and McGurk conditions. As such, we used audio responses to McGurk stimuli to define correct responses to congruent stimuli since the participants were instructed to respond based on the audio (i.e., what they heard). In this case, increased audio-dominant responses would indicate decreased visual-dominant responses.

**Table 1.**

Means and standard deviations of correct response rate in the various stimulus conditions, and correlations between correct response rate and Autism-Spectrum Quotient (AQ)

	Mean	SD	Pearson's correlation coefficient with AQ		
			<i>r</i>	CI	BF <sub>10</sub>
Congruent condition	0.91	0.073	−0.09	−0.43–0.28	0.25
McGurk condition					
audio response ('apa')	0.48	0.223	0.36*	0.01–0.63	1.42
fused response ('ata')	0.45	0.151	−0.33 <sup>†</sup>	−0.61–0.02	1.10
visual response ('aka')	0.07	0.136	−0.21	−0.53–0.15	0.42
Visual condition	0.66	0.109	−0.04	−0.39–0.32	0.23
Auditory condition	0.89	0.089	−0.1	−0.44–0.26	0.26

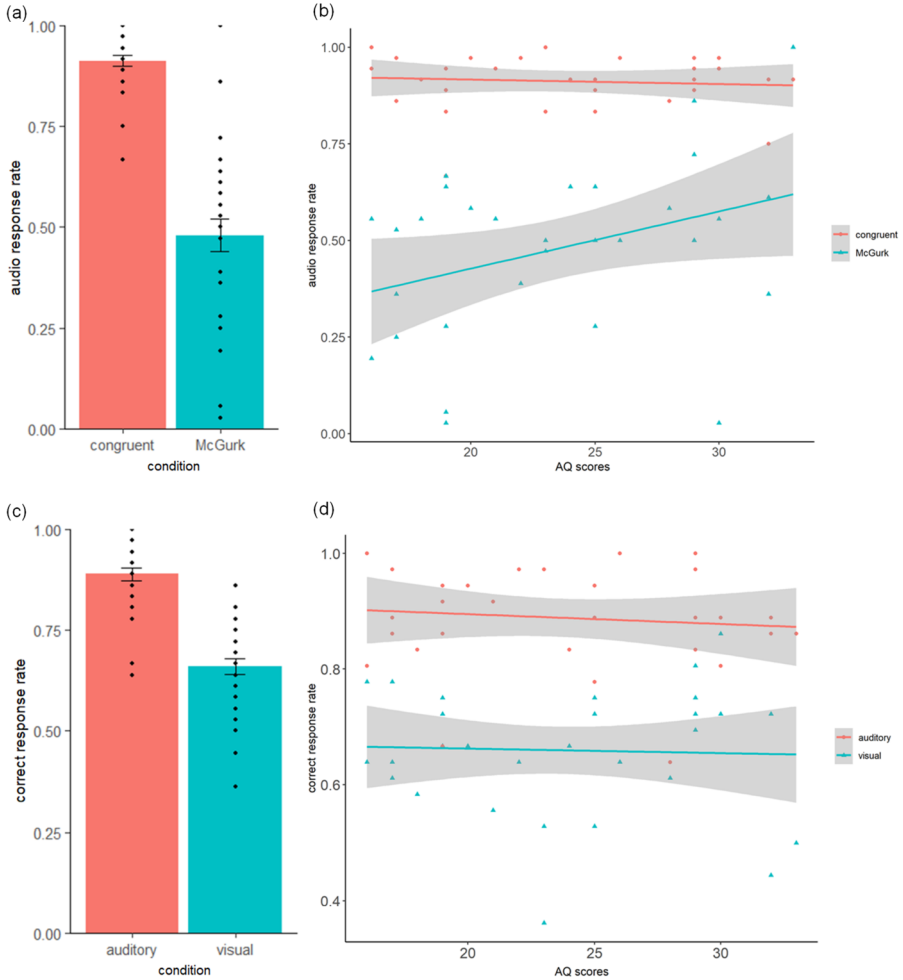
In the McGurk condition, there were three possible responses: audio ('apa' response), fused ('ata' response), and visual ('aka' response). CI indicates a 95 % confidence interval for each correlation. BF<sub>10</sub> indicates Bayes factors for each correlation. <sup>†</sup>,  $p < 0.10$ ; \*,  $p < 0.05$ .

### 3. Results

We calculated mean correct (audio) responses for congruent and McGurk conditions and mean correct responses for auditory and visual conditions (Table 1). Participants' mean AQ score was 23.6 (SD = 5.37), with their scores ranging from 16 to 33. The AQ results were normally distributed in the current sample (Kolmogorov–Smirnov test:  $D = 0.16$ ,  $p = 0.42$ ). In addition, Kolmogorov–Smirnov tests revealed that the data of each condition were normally distributed (visual:  $D = 0.14$ ,  $p = 0.61$ ; auditory:  $D = 0.12$ ,  $p = 0.74$ ; congruent:  $D = 0.20$ ,  $p = 0.16$ ; McGurk, for auditory responses:  $D = 0.18$ ,  $p = 0.27$ ). Then, we examined the relationships between task performance and AQ (Figs 2b and 2d). In the correlation analysis, AQ was significantly and positively correlated with audio ('apa') responses in the McGurk condition [ $r = 0.36$ ,  $p = 0.049$ , 95% confidence interval (CI) = 0.01–0.63, Bayes factor (BF<sub>10</sub>) = 1.42], while the correlation was not significant in the audiovisual congruent condition ( $r = -0.09$ ,  $p = 0.648$ , 95% CI = −0.43–0.28, BF<sub>10</sub> = 0.25). Furthermore, there was no significant correlation of AQ with correct responses in the auditory ( $r = -0.10$ ,  $p = 0.581$ , 95% CI = −0.44–0.26, BF<sub>10</sub> = 0.26) nor in the visual condition ( $r = -0.04$ ,  $p = 0.837$ , 95% CI = −0.39–0.32, BF<sub>10</sub> = 0.23).

In order to test whether AQ influenced audiovisual speech integration, we conducted a multiple regression analysis with correct (audio) response rate as the dependent variable, and stimulus type (congruent and McGurk-type), AQ, and their interaction as predictor variables. We found that stimulus type ( $\beta = -0.798$ ,  $t = 10.71$ ,  $p < 0.01$ ) was a significant predictor of the correct audio response rate. This indicates the occurrence of the McGurk effect in the





**Figure 2.** Summary of the results of the experiment: (a) mean correct (audio) responses in congruent and McGurk conditions; (b) relationship between Autism Spectrum Quotient (AQ) and correct (audio) responses in congruent and McGurk conditions; (c) mean correct responses in auditory and visual conditions; and (d) relationship between AQ and correct responses in auditory and visual conditions.

incongruent stimuli, demonstrating that the silhouette of the profile of facial speech also had a visual influence on audiovisual speech perception. The predictive effect of AQ was not significant ( $\beta = -0.023$ ,  $t = 0.21$ ,  $p = 0.83$ ). Furthermore, the interaction between stimulus type and AQ ( $\beta = 0.220$ ,  $t = 2.09$ ,  $p = 0.041$ ) was a significant predictor of correct audio response rate. A simple slope test showed that AQ predicted the audio response rate in the

incongruent condition ( $\beta = 0.015$ ,  $t = 2.74$ ,  $p = 0.008$ ), but not in the congruent condition ( $\beta = -0.001$ ,  $t = 0.23$ ,  $p = 0.822$ ). This indicates a weaker influence of visual speech on audiovisual speech perception in participants with higher AQ than in those with lower AQ. However, such results should be carefully interpreted since the variance in the congruent condition was smaller than that in the McGurk condition, although no ceiling effects were observed in both conditions.

#### 4. Discussion

The current study tested the relationship between autistic traits and the influence of visual speech on audiovisual speech perception of Rubin's vase-type stimuli, which contained silhouettes of a speaker's profile articulating syllables. As an index of the magnitude of visual influence, we used audio responses to the McGurk stimuli, as in previous studies (Ujiie *et al.*, 2015, 2018). More audio-dominant responses corresponded to fewer fused and visual responses, namely visual influence on speech was decreased in the former case. The results demonstrated that the audio response rate in the McGurk condition was significantly higher in individuals with higher AQ scores than in those with lower AQ scores. This result is consistent with previous results (e.g., Ujiie *et al.*, 2015, 2018), although the correlation ( $r = 0.36$ ,  $BF_{10} = 1.42$ ) needs to be interpreted as relatively weak (Cohen, 1988; Kass and Raftery, 1995). Thus, even for Rubin's vase speech stimuli, the influence of visual speech (i.e., the McGurk effect) weakened with increasing autistic traits. In contrast, autistic traits were not correlated with the accuracy of lip-reading and listening. The results also indicated that autistic traits were not significantly correlated with the accuracy of recognizing audiovisual congruent speech, although this lack of relationship may have been due to the small response variance in the audiovisual congruent condition.

##### 4.1. *A Weak McGurk Effect in Autism Spectrum Disorder: the Effect of Structural Face Processing*

The present study suggested that a weaker McGurk effect in autistic individuals may be associated with the processing of kinematic information when structural face information was weakened by using silhouettes. The silhouette stimulus contained kinematic information whereas the facial components and textures were removed so that the structural facial information was degraded. Taken together, the results implied that the kinematic information would primarily contribute to the weaker McGurk effect in individuals with a high AQ. It is noteworthy that we did not exclude the possibility that the structural information, if available, may also contribute to the weaker McGurk effect, since kinematic and structural information are related to the occurrence of

McGurk effect (e.g., Rosenblum and Saldaña, 1996; Rosenblum *et al.*, 2000). The results also suggested that the weaker visual influence in participants with high versus low AQs did not arise from a difference in the observer's conscious perception (e.g., Munhall *et al.*, 2009; Peterson *et al.*, 1991). Munhall *et al.* (2009) showed that Rubin's vase-type visual speech led to the McGurk effect when an observer noticed that visual speech was conveyed by the silhouettes of the profiles of two speakers, whereas in the experiment, we instructed the participants to observe the stimuli as movies of articulated syllables. As a result, a reduced rate of correct audio responses was found in the McGurk condition compared with the congruent condition, and this rate was comparable between participants with higher and lower AQ. In addition, a previous study found no difference between the performance of individuals with and without a diagnosis of ASD on the Rubin's vase task of recognizing a face in a Rubin's image (Akechi *et al.*, 2014). Therefore, we hypothesized that weaker visual influence in Rubin's vase-type visual speech would not derive from differences in the ability to detect faces in Rubin vase images.

#### 4.2. *A Weak McGurk Effect in ASD: the Effect of Kinematic Facial Processing*

However, our results did not directly indicate that the weak McGurk effect was caused by atypical processing of kinematic facial information. Our results showed that the accuracy of lip-reading was comparable across all participants regardless of individuals' AQ, although a weaker McGurk effect was observed in individuals with higher autistic traits. Indeed, several studies have reported that a weaker McGurk effect is not due to poor lip-reading abilities (e.g., de Gelder *et al.*, 1991; Saalasti *et al.*, 2012). Although autistic children show a delay in the development of lip-reading ability (Taylor *et al.*, 2010), autistic adults achieve the same level as TD adults (e.g., Saalasti *et al.*, 2012). Regarding the lip-reading ability, our results did not provide evidence whether audiovisual gains decreased along with AQ scores, since we did not manipulate the auditory noise level. The magnitude of audiovisual gains can be calculated by subtracting accuracies for auditory speech from those for audiovisual speech for speech-in-noise performance (e.g., Foxe *et al.*, 2013). If autistic people have deficits in processing kinematic facial information, lower audiovisual gains would be observed. Although our results showed no significant correlation between AQ scores and accuracies for audiovisual congruent speech, this result may be limited because of the lower variances in this condition compared to the McGurk condition. Future investigations should address this issue by using speech recognition tasks with several noise levels.

### 4.3. Another Possible Explanation: Weak Central Coherence

The weaker McGurk effect in ASD may be associated with a perceptual bias in the unification of multisensory information, based on weak central coherence (Frith, 1991; Happé and Frith, 2006). A detail-focused processing style in ASD has been proposed to explain weak central coherence, which suggests that autistic individuals prefer local information to global contextual information. Originally, this model assumed that local bias appears in perceptual processing in unisensory domains, such as in the embedded figure test or the Navon-type global local naming task (Frith, 1991). Recently, it has been assumed to appear also in multisensory processing (e.g., Happé and Frith, 2006), that is, detail-focused processing leads to difficulty in the ‘grouping of stimuli’, which may cause a weaker reliance on visual input during multisensory processing. Indeed, by demonstrating a close link between poor acuity of audiovisual temporal binding and a weaker McGurk effect in autistic individuals, Stevenson *et al.* (2014) suggested that common deficits emerge in the integration processes at both lower and higher levels.

### 4.4. Limitations

Our study has some limitations. First, our results may not be generalizable to everyone with a diagnosis of ASD., since our study used an analogous design to study ASD symptoms in a typically developing sample, namely an indirect investigation in a population that was not diagnosed with ASD. However, our results are generalizable when interpreted on the basis of the dimensional model of ASD (e.g., Baron-Cohen *et al.*, 2001a; Frith, 1991). This model assumes that the distribution of AQ scores forms a continuum across non-clinical and clinical populations, and people with ASD are assumed to represent extremes of AQ scores (e.g., Baron-Cohen, 2002; Baron-Cohen *et al.*, 2001a; Wakabayashi *et al.*, 2006; Woodbury-Smith *et al.*, 2005). Indeed, numerous studies have shown that people with high AQs but no diagnosis of ASD exhibit similar tendencies in perceptual and cognitive tasks to people diagnosed with ASD (Reed *et al.*, 2011; Stewart and Ota, 2008; Woodbury-Smith *et al.*, 2005), including for the McGurk effect (e.g., Ujiie and Wakabayashi, 2019; Ujiie *et al.*, 2015, 2018; van Laarhoven *et al.*, 2019). Further, our sample included people with higher AQs than the cutoff point between clinical and non-clinical populations (cutoff point of 27; Woodbury-Smith *et al.*, 2005), who are considered to exhibit a broader autism phenotype (e.g., Austin, 2005; Hasegawa *et al.*, 2015). Therefore, the results from analogous study can contribute to understanding perceptual and behavioral tendencies in populations diagnosed with ASD, albeit indirectly.

## 5. Conclusion

In conclusion, our results suggest that susceptibility to the McGurk effect as assessed with Rubin's vase-type stimuli lies on a continuum consistent with self-reported autism features in the neurotypical population. This may be indicative of difficulties with audiovisual integration, rather than difficulties with unisensory processing.

### *Acknowledgements*

We would like to thank all the study participants.

### *Funding Information*

This study was supported by a Grant-in-Aid for Early-Career Scientists (Grant No. 19K20650), the Promotion of Joint International Research (Grant No. 20KK0054), and a Grant-in-Aid for JSPS Fellows (Grant No. 19J00722). This work was also supported by a Grant-in-Aid for Scientific Research on Innovative Areas 'Construction of the Face-Body Studies in Transcultural Conditions' (Grant No. 17H06342).

### *Conflicts of Interest*

The authors have no competing financial interests to declare.

### *Author Contributions*

All authors contributed to the study design. YU performed testing, data collection, data analysis, and interpretation under the supervision of KT. YU drafted the manuscript and KT provided critical revisions. All authors approved the final version of the manuscript for submission.

## References

- Akechi, H., Kikuchi, Y., Tojo, Y., Osanai, H. and Hasegawa, T. (2014). Neural and behavioural responses to face-likeness of objects in adolescents with autism spectrum disorder, *Sci. Rep.* **4**, 3874. DOI:10.1038/srep03874.
- American Psychiatric Association (2013). *Diagnostic and Statistical Manual of Mental Disorders*, 5th edn. American Psychiatric Association, Washington, DC, USA.
- Austin, E. J. (2005). Personality correlates of the broader autism phenotype as assessed by the Autism Spectrum Quotient (AQ), *Pers. Individ. Dif.* **38**, 451–460. DOI:10.1016/j.paid.2004.04.022.
- Baron-Cohen, S. (1995). *Mindblindness: an Essay on Autism and Theory of Mind*. MIT Press/Bradford Books, Cambridge, MA, USA.
- Baron-Cohen, S. (2002). The extreme male brain theory of autism, *Trends Cogn. Sci.* **6**, 248–254. DOI:10.1016/s1364-6613(02)01904-6.

- Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J. and Clubley, E. (2001a). 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.
- Baron-Cohen, S., Wheelwright, S., Hill, J., Raste, Y. and Plumb, I. (2001b). 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.
- Beauchamp, M. S., Argall, B. D., Bodurka, J., Duyn, J. H. and Martin, A. (2004). Unraveling multisensory integration: patchy organization within human STS multisensory cortex, *Nat. Neurosci.* **7**, 1190–1192. DOI:10.1038/nn1333.
- Bebko, J. M., Schroeder, J. H. and Weiss, J. A. (2014). The McGurk effect in children with autism and Asperger syndrome, *Autism Res.* **7**, 50–59. DOI:10.1002/aur.1343.
- 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.
- Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*, 2nd edn.. Lawrence Erlbaum Associates, Hillsdale, NJ, USA.
- 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.
- 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.
- Eskelund, 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.
- Foxe, J. J., Molholm, S., Del Bene, V. A., Frey, H.-P., Russo, N. N., Blanco, D., Saint-Amour, D. and Ross, L. A. (2015). Severe multisensory speech integration deficits in high-functioning school-aged children with Autism Spectrum Disorder (ASD) and their resolution during early adolescence, *Cereb. Cortex* **25**, 298–312. DOI:10.1093/cercor/bht213.
- Frith, U. (1991). *Autism and Asperger’s Syndrome*. Cambridge University Press, Cambridge, UK. DOI:10.1017/CBO9780511526770.
- 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.
- Hasegawa, C., Kikuchi, M., Yoshimura, Y., Hiraishi, H., Munesue, T., Nakatani, H., Higashida, H., Asada, M., Oi, M. and Minabe, Y. (2015). Broader autism phenotype in mothers predicts social responsiveness in young children with autism spectrum disorders, *Psychiatry Clin. Neurosci.* **69**, 136–144. DOI:10.1111/pcn.12210.
- Hasson, U., Hendler, T., Ben Bashat, D. and Malach, R. (2001). Vase or face? A neural correlate of shape-selective grouping processes in the human brain, *J. Cogn. Neurosci.* **13**, 744–753. DOI:10.1162/08989290152541412.
- Hesselmann, G., Kell, C. A., Eger, E. and Kleinschmidt, A. (2008). Spontaneous local variations in ongoing neural activity bias perceptual decisions, *Proc. Natl Acad. Sci. U. S. A.* **105**, 10984–10989. DOI:10.1073/pnas.0712043105.

- 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. <https://doi.org/10.1007/s10803-008-0538-x>.
- Iarocci, G., Rombough, A., Yager, J., Weeks, D. J. 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 facial 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., Mccotter, M. V. and Thomas, S. M. (2000). Visual and audiovisual speech perception with color and gray-scale facial images, *Percept. Psychophys.* **62**, 1394–1404. DOI:10.3758/bf03212141.
- Kass, R. E. and Raftery, A. E. (1995). Bayes factors, *J. Am. Stat. Assoc.* **90**, 773–795. DOI:10.1080/01621459.1995.10476572.
- Kättyri, 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. <https://doi.org/10.1016/j.rasd.2009.09.015>.
- Kikuchi, Y., Senju, A., Hasegawa, T., Tojo, Y. and Osanai, H. (2013). The effect of spatial frequency and face inversion on facial expression processing in children with autism spectrum disorder, *Jpn. Psychol. Res.* **55**, 118–130. DOI:10.1111/jpr.12000.
- Koldewyn, K., Whitney, D. and Rivera, S. M. (2011). Neural correlates of coherent and biological motion perception in autism, *Dev. Sci.* **14**, 1075–1088. DOI:10.1111/j.1467-7687.2011.01058.x.
- 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.
- Lieberman, A. M. and Mattingly, I. G. (1985). The motor theory of speech perception revised, *Cognition* **21**, 1–36. DOI:10.1016/0010-0277(85)90021-6.
- MacDonald, J. (2018). Hearing lips and seeing voices: the origins and development of the ‘McGurk effect’ and reflections on audio–visual speech perception over the last 40 years, *Multisens. Res.* **31**, 7–18. DOI:10.1163/22134808-00002548.
- McGurk, H. and MacDonald, J. W. (1976). Hearing lips and seeing voices, *Nature* **264**, 746–748. DOI:10.1038/264746a0.
- Munhall, K. G., ten Hove, M. W., Brammer, M. and Paré, M. (2009). Audiovisual integration of speech in a bistable illusion, *Curr. Biol.* **19**, 735–739. DOI:10.1016/j.cub.2009.03.019.
- O’Brien, J., Spencer, J., Girges, C., Johnston, A. and Hill, H. (2014). Impaired perception of facial motion in autism spectrum disorder, *PLoS ONE* **9**, e102173. DOI:10.1371/journal.pone.0102173.

- Peterson, M. A., Harvey, E. M. and Weidenbacher, H. J. (1991). Shape recognition contributions to figure-ground reversal: which route counts? *J. Exp. Psychol. Hum. Percept. Perform.* **17**, 1075–1089. DOI:10.1037//0096-1523.17.4.1075.
- 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.
- Rosenblum, L. D. (2008). Speech perception as a multimodal phenomenon, *Curr Dir. Psychol. Sci.* **17**, 405–409. DOI:10.1111/j.1467-8721.2008.00615.x.
- Rosenblum, L. D. and Saldaña, H. M. (1996). An audiovisual test of kinematic primitives for visual speech perception, *J. Exp. Psychol. Hum. Percept. Perform.* **22**, 318–331. DOI:10.1037//0096-1523.22.2.318.
- Rosenblum, L. D., Yakel, D. A. and Green, K. P. (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.
- 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.
- Rouse, H., Donnelly, N., Hadwin, J. A. and Brown, T. (2004). Do children with autism perceive second-order relational features? The case of the Thatcher illusion, *J. Child Psychol. Psychiatry* **45**, 1246–1257. DOI:10.1111/j.1469-7610.2004.00317.x.
- Saalisti, 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.
- Stevenson, R. A., Siemann, J. K., Schneider, B. C., Eberly, H. E., Woynaroski, T. G., Camarata, S. M. and Wallace, M. T. (2014). Multisensory temporal integration in autism spectrum disorders, *J. Neurosci.* **34**, 691–697. DOI:10.1523/JNEUROSCI.3615-13.2014.
- 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.
- Sumby, 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.
- Tanaka, J. W. and Sung, A. (2016). The “eye avoidance” hypothesis of autism face processing, *J. Autism Dev. Disord.* **46**, 1538–1552. DOI:10.1007/s10803-013-1976-7.
- 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.
- Teunisse, J. P. and de Gelder, B. (2001). Impaired categorical perception of facial expressions in high-functioning adolescents with autism, *Child Neuropsychol.* **7**, 1–14. DOI:10.1076/chin.7.1.1.3150.
- Tiippana, K. (2014). What is the McGurk effect? *Front. Psychol.* **5**, 725. DOI:10.3389/fpsyg.2014.00725.
- Ujiié, 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.
- Ujiié, Y. and Wakabayashi, A. (2019). Intact lip-reading but weaker McGurk effect in individuals with high autistic traits, *Int. J. Dev. Disabil.* DOI:10.1080/20473869.2019.1699350.



- 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.
- Ujiie, Y., Asai, T. and Wakabayashi, A. (2018). Individual differences and the effect of face configuration information in the McGurk effect, *Exp. Brain Res.* **236**, 973–984. DOI:10.1007/s00221-018-5188-4.
- van Laarhoven, T., Stekelenburg, J. J. and Vroomen, J. (2019). Increased sub-clinical levels of autistic traits are associated with reduced multisensory integration of audiovisual speech, *Sci. Rep.* **9**, 9535. DOI:10.1038/s41598-019-46084-0.
- Wakabayashi, A., Baron-Cohen, S., Wheelwright, S. and Tojo, Y. (2006). The Autism-Spectrum Quotient (AQ) in Japan: a cross-cultural 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. <https://doi.org/10.1016/j.ridd.2004.01.008>.
- Woodbury-Smith, M. R., Robinson, J., Wheelwright, S. and Baron-Cohen, S. (2005). Screening adults for Asperger syndrome using the AQ: a preliminary study of its diagnostic validity in clinical practice, *J. Autism Dev. Disord.* **35**, 331–335. <https://doi.org/10.1007/s10803-005-3300-7>.
- Woynaroski, 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. <https://doi.org/10.1007/s10803-013-1836-5>.
- Zhang, J., Meng, Y., He, J., Xiang, Y., Wu, C., Wang, S. and Yuan, Z. (2019). McGurk effect by individuals with autism spectrum disorder and typically developing controls: a systematic review and meta-analysis, *J. Autism Dev. Disord.* **49**, 34–43. DOI:10.1007/s10803-018-3680-0.