Received: 9 February 2021 | Accepted: 30 July 2021

DOI: 10.1002/aur.2598

RESEARCH ARTICLE

Face- ie ing pa ern predic a dio i al peech in egra ion in a i ic children

Sh an Feng[1,2] Hao ang L [3,4] Qiandong Wang[5] Tianbi Li[2] Jing Fang[6] Lihan Chen[2] Li Yi[2,7]

[1]Institute for Applied Linguistics, School of Foreign Languages, Central South University, Changsha, Hunan, China

[2]School of Psychological and Cognitive Sciences and Beijing Key Laboratory of Behavior and Mental Health, Peking University, Beijing, China

[3]Academy for Advanced Interdisciplinary Studies, Peking University, Beijing, China

[4]Peking-Tsinghua Center for Life Sciences, Peking University, Beijing, China

[5]Beijing Key Laboratory of Applied Experimental Psychology, National Demonstration Center for Experimental Psychology Education, Faculty of Psychology, Beijing Normal University, Beijing, China

[6]Qingdao Autism Research Institute, Qingdao, China

[7]IDG/McGovern Institute for Brain Research at PKU, Peking University, Beijing, China

Corre pondence
Lihan Chen and Li Yi, Peking University, 5 Yiheyuan Road, Beijing, 100871, China.
Email: clh@pku.edu.cn; yilipku@pku.edu.cn

F nding informa ion
Guangdong Key Project in "Development of new tools for diagnosis and treatment of Autism", Grant/Award Number: 2018B030335001; National Natural Science Foundation of China, Grant/Award Numbers: 31871116, 31861133012; Philosophy and Social Science Foundation of Hunan Province, Grant/Award Number: 19YBQ109

Ab rac

Autistic children show audiovisual speech integration deficits, though the underlying mechanisms remain unclear. The present study examined how audiovisual speech integration deficits in autistic children could be affected by their looking patterns. We measured audiovisual speech integration in 26 autistic children and 26 typically developing (TD) children (4- to 7-year-old) employing the McGurk task (a videotaped speaker uttering phonemes with her eyes open or closed) and tracked their eye movements. We found that, compared with TD children, autistic children showed weaker audiovisual speech integration (i.e., the McGurk effect) in the open-eyes condition and similar audiovisual speech integration in the closed-eyes condition. Autistic children viewed the speaker's mouth less in non-McGurk trials than in McGurk trials in both conditions. Importantly, autistic children's weaker audiovisual speech integration could be predicted by their reduced mouth-looking time. The present study indicated that atypical face-viewing patterns could serve as one of the cognitive mechanisms of audiovisual speech integration deficits in autistic children.

La S mmar

McGurk effect occurs when the visual part of a phoneme (e.g., "ga") and the auditory part of another phoneme (e.g., "ba") uttered by a speaker were integrated into a fused perception (e.g., "da"). The present study examined how McGurk effect in autistic children could be affected by their looking patterns for the speaker's face. We found that less looking time for the speaker's mouth in autistic children could predict weaker McGurk effect. As McGurk effect manifests audiovisual speech integration, our findings imply that we could improve audiovisual speech integration in autistic children by directing them to look at the speaker's mouth in future intervention.

KEYWORDS
atypical face-viewing patterns, audiovisual speech integration, autistic children, eye movements, McGurk effect

INTRODUCTION

Autism spectrum disorder (ASD) is a neurodevelopmental disorder characterized by impaired social interaction and communication, as well as restricted and repetitive patterns of behavior (DSM-5; American Psychiatric Association, 2013). Autistic children have been reported to have weaker audiovisual speech integration than typically developing (TD) children (see a meta-analysis by Zhang et al., 2019). Audiovisual speech integration plays important roles in language and social development, especially in the early stage of life, by affecting speech comprehension and language

[Correction added on 01 September 2021, after first online publication: The grant numbers in funding information have been revised in this version.]

processing during social interactions (Gervain & Mehler, 2010; Werker & Gervain, 2013). Audiovisual speech integration deficits in autistic children could interfere with their social communications (Feldman et al., 2018).

Audiovisual speech integration is often measured by the McGurk effect (McGurk & MacDonald, 1976). The McGurk effect occurs when dubbing the acoustic part of a phoneme (e.g., "ba") onto the visual part of another phoneme (e.g., "ga"), resulting in a fused perception of a third phoneme (e.g., "da"; McGurk & MacDonald, 1976). Most studies using the McGurk paradigm to investigate audiovisual speech integration found that autistic children showed weaker audiovisual speech integration (i.e., a weaker McGurk effect) than TD children (Bebko et al., 2014; de Gelder et al., 1991; Irwin et al., 2011; Mongillo et al., 2008; Stevenson et al., 2014). Other studies found that autistic children and TD children showed similar audiovisual speech integration (i.e., McGurk effect; Iarocci et al., 2010; Woynaroski et al., 2013). These inconsistencies between previous studies could be accounted for by participants' characteristics and the different criteria for the McGurk effect measurement (Zhang et al., 2019). The weaker audiovisual speech integration in autistic children might be derived from their cognitive impairments, such as their weaker tendency to combine stimulus parts into a coherent one as proposed by the weak central coherence theory (Baum et al., 2015; Happé & Frith, 2006), their multisensory temporal processing deficits (Stevenson et al., 2014), and their atypical face viewing patterns (Bebko et al., 2014; Irwin et al., 2011). The present study aims to discover the association between atypical eye gaze to faces in autistic children and their audiovisual speech integration.

Face viewing plays a fundamental role in social communications (Fort et al., 2013; Jack & Schyns, 2015; Pascalis et al., 2014; Sumby & Pollack, 1954). Face viewing provides us with information about those with whom we interact, such as individuals with certain gender, age, identity, race, emotion, personality, and even social status (Jack & Schyns, 2015; Pascalis et al., 2014). In addition, face viewing is crucial for speech perception; for example, it improves core word recognition and enhances speech perception in noisy environments (Fort et al., 2013; Sumby & Pollack, 1954). However, autistic individuals exhibited atypical face viewing patterns: they displayed diminished looking time at the core features (i.e., eyes, mouth, and nose) compared with TD individuals (Pelphrey et al., 2002). In particular, autistic children viewed the mouth less than TD children when they performed the face recognition task (Chawarska & Shic, 2009), scanned emotional faces (de Wit et al., 2008), or viewed videos of talking people (Nakano et al., 2010). Moreover, previous studies have found that audiovisual speech integration is associated with face viewing patterns in TD individuals; that is, the McGurk effect increases as the mouth-looking time increases (Gurler et al., 2015). Despite the evidence for deficits in both face viewing patterns and audiovisual speech integration in autistic children, the relationship between them remains unclear. Thus, in the present study, we investigated the relationship between atypical face viewing patterns and audiovisual speech integration deficits (i.e., the weaker McGurk effect) in autistic children.

To address the above research question, we employed the McGurk effect to measure audiovisual speech integration in autistic children. In addition to the typical McGurk setting in which the speaker uttered phonemes with her eyes open (open-eyes condition), we set a closed-eyes condition in which the speaker uttered phonemes with her eyes closed. This manipulation could have different effects on audiovisual speech integration for autistic children and TD children. On the one hand, for autistic children, the closed-eyes condition in the present study would help eliminate the threatening effect of the speaker's direct gaze as proposed by the eye avoidance hypothesis and divert the children's attention to the speaker's mouth movements (i.e., increase the mouth-looking time), hence enhancing their audiovisual speech integration (Tanaka & Sung, 2016). The eye avoidance hypothesis held that autistic individuals avoid viewing the eyes as they perceived eyes as socially threatening. On the other hand, the closed-eyes condition would weaken audiovisual speech integration in TD children, given that the closed-eyes manipulation eliminated the eyes' information which is crucial for speech perception, or that the closed-eyes face might induce an extra cognitive processing (Davidhizar, 1992; Happé & Frith, 2006; Holler et al., 2014). During the experiment, children were asked to report what the speaker said after viewing videos of a speaker uttering phonemes, while their eye movements were tracked during the experiment. Based on previous literature, we hypothesized that (a) autistic children would display reduced audiovisual speech integration compared with TD children, (b) autistic children would demonstrate an atypical face viewing pattern, and (c) the atypical face viewing pattern in autistic children would predict their reduced audiovisual speech integration (i.e., McGurk effect). We also hypothesize that the closed-eyes condition would improve audiovisual speech integration in autistic children (by eliminating the eyes' threatening information) but weaken audiovisual speech integration in TD children (by eliminating the eyes' information, or by inducing an extra cognitive processing).

METHOD

Par icipan

Twenty-six TD boys and 26 autistic boys participated in the present study. We recruited TD children from a kindergarten and an elementary school and autistic children from a school specializing in autistic children in China. All autistic children had received a diagnosis of ASD based on the criteria of the Diagnostic and Statistical Manual of Mental Disorders (DSM-V; American

Psychiatric Association, 2013). We further confirmed the ASD diagnosis according to the Chinese version of the Autism Spectrum Quotient: Children's Version (AQ-Child; Auyeung et al., 2008). The two groups were matched in age, Full Scale Intelligence Quotient (FSIQ), and Verbal Comprehension Index (VCI; see Table 1 for detailed information). Age ranged from 4.84 to 7.53 in the autistic group, and from 4.40 to 7.08 in the TD group. FSIQ scores ranged from to 83 to 126 in the autistic group and from 86 to 136 in the TD group. VCI scores ranged from 68 to 132 in the autistic group, and from 81 to 126 in the TD group. FSIQ and VCI were measured by the Chinese version of abbreviated Wechsler Preschool and Primary Scale of Intelligence-Fourth Edition (WPPSI-IV; Wechsler, 2014). Before the study, children and their parents were informed that parents needed to complete a questionnaire (AQ-child) and that children needed to first participate in an IQ test and then in an experiment. In the experiment, children needed to report what the speaker said and their eye movements would be tracked. After the experiment, children would receive a gift. Children and their parents were also told that children were voluntary in participating the present study and that children's information would be treated confidentially. We obtained oral consent from all children and written consent from their parents before the experiment.

Stimuli

The present study employed the McGurk effect to measure participants' audiovisual speech integration (McGurk & MacDonald, 1976) and used phonemes similar to those used in Stevenson et al. (2014). The experiment included two conditions: open- and closed-eyes conditions. In these two conditions, a female speaker articulated phonemes with her eyes open or closed. Each condition included congruent stimuli and incongruent stimuli. The congruent stimuli were the original videos of the speaker articulating "ba," "ga," "pa," and "ka." The incongruent stimuli were modified from the original videos by dubbing the visual "ga" onto the auditory "ba" ("AbVg": auditory "ba" + visual "ga") or by dubbing the visual "ka" onto the auditory "pa" ("ApVk": auditory "pa" + visual "ka") using Adobe Premiere Software Pro CS 6.0. These modifications could evoke the McGurk illusory percept of "da" and "ta" respectively. In the formal experiment, we used "ba," "ga," and "AbVg"; in the practice session, we used "pa," "ka," and "ApVk."

All the stimuli were videos with a resolution of 1280 × 720 pixels and a frame rate of 25 frames/s. Each stimulus lasted approximately 2.10 s. All videos began and ended with the speaker's still face. In the open-eyes condition, the speaker opened her mouth gradually from 0.1 s to 1 s and closed her mouth gradually from 1 s to 2.10 s. In the closed-eyes condition, the speaker opened her mouth gradually from 0.25 s to 1.3 s and closed her mouth gradually from 1.3 s to 2.10 s. We obtained written consent from the female speaker to use these videos in the experiment and publications.

Procedure

The experiment was conducted in a quiet room. Participants were seated approximately 60 cm from the display screen. The display screen was a 21.5-inch Dell display screen with a resolution of 1920 × 1080 pixels. In the center of the screen, the stimuli were presented with Psychotoolbox (Brainard, 1997; Kleiner et al., 2007; Pelli, 1997) and MATLAB (The MathWorks, Natick, MA). Sounds were presented through two speakers located on the two sides of the screen. We recorded children's eye movements using a Tobii X 120 eye tracker with a sampling rate of 120 Hz.

Participants first completed the practice session to become familiar with the task, in which they were required to verbally report what the speaker said. At the beginning of the formal experiment, we calibrated participants' eye movements by a five-point calibration procedure. The calibration was accepted only when all five points showed a good fit, with error vectors smaller than 0.5 degrees of visual angle. The formal experiment included the open- and closed-eyes conditions. Each condition consisted of 10 trials of congruent "ba," 10 trials of congruent "ga," and 12 trials of incongruent "AbVg" (auditory "ba" + visual "ga"). The 32 trials in each condition were randomly presented. For each trial, a fixation was first displayed on the screen for 1000 ms, and participants were asked to look at it. Then, a black screen was displayed for 800 ms. Next, the stimulus was presented. Finally, a black screen was shown until participants responded. The procedure of a sample trial is shown in Figure 1. The experimenter recorded participants' responses by pressing the corresponding button on the

TABLE 1 Characteristics of the TD children and the ASD children

	N	Male/female	Mean age in year (*SD*)	Mean FSIQ[a] (*SD*)	Mean VCI[a] (*SD*)
ASD	26	26/0	6.07 (0.81)	101.50 (11.24)	101.92 (16.06)
TD	26	26/0	5.76 (0.88)	104.54 (13.10)	102.85 (13.62)
ASD vs. TD (*t* value)[b]	N/A	N/A	1.30	−0.90	−0.22

[a]FSIQ and VCI were measured by the Chinese version of abbreviated Wechsler Preschool and Primary Scale of Intelligence-Fourth Edition (WPPSI-IV).
[b]All $ps > 0.05$.

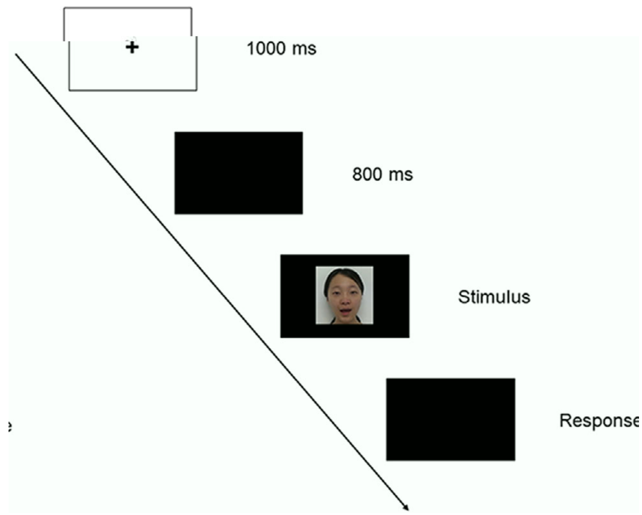

FIGURE 1 Procedure of a sample trial. Each trial began with fixation at the center of the screen for 1000 ms, followed by a black screen for 800 ms. Then, the stimulus was displayed. Finally, a black screen was presented, and participants were asked to respond

keyboard (i.e., pressed "b," "d," and "g" when children responded "ba," "da," and "ga," respectively). The experiment lasted approximately 25 minutes.

Da a anal i

Behavioral data analysis

In both conditions, for the congruent trials, only responses that matched the stimuli were coded as correct. For the incongruent trials, participants made one of the four kinds of responses: auditory response "ba," visual response "ga," fused response "da" (McGurk response), and other responses, such as /a/. We analyzed the data using nonparametric statistical analysis (i.e., Mann–Whitney *U* test, Wilcoxon signed rank test, or permutation ANOVA) as the normal distribution assumption of the corresponding parametric tests was violated.

Eye movement data analysis

We defined five areas of interest (AOIs) in each condition: the whole face, the eyes (i.e., left eye and right eye), the mouth, the nose, and the other area (i.e., the area on the face except for the eyes, nose, and mouth; see Figure 2). We extracted fixations from the raw gaze data according to the definition specified by Tobii (I-VT fixation filter; Olsen, 2012). Specifically, the minimum fixation duration was set at 60 ms within a velocity of 30 deg/s. We computed the participant's *total looking time* on each AOI by summing the total durations of all fixations within the AOI. We also calculated the

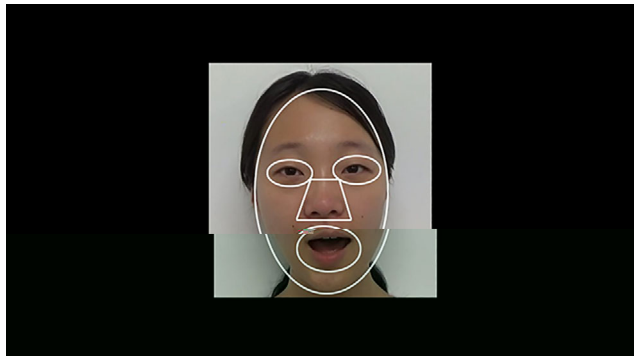

FIGURE 2 Sample area of interest (AOI) used in the eye movement data analysis. The AOIs include the whole face, eyes, mouth, nose, and the other area. See the text for details

participant's *proportional looking time* on the eye area, the mouth area, the nose area, and other areas by dividing the total time looking at each of the areas by the total time looking at the whole face.

Considering that *mouth-looking time* was crucial for evoking audiovisual speech integration (i.e., the McGurk effect; Gurler et al., 2015) and that mouth-looking behavior possibly changed over time, we further explored the temporal course of mouth-looking time. In particular, we examined (a) how the mouth-looking time changed over time and (b) whether it differed between groups and between McGurk trials (trials evoking the McGurk effect) and non-McGurk trials (trials not evoking the McGurk effect) within each group (Gurler et al., 2015). We also analyzed the temporal course of looking time for other AOIs (the eyes, the nose, the other area, and the area away from the face), and explored whether it differed between McGurk trials and non-McGurk trials within each group in both conditions. The temporal-course analysis was based on the moving-average approach (Dankner et al., 2017). We segmented the data of each trial into epochs of 1/6 s (20-sample data) with a step of 1/120 s (one-sample data). We then calculated the mouth-looking time in each epoch, which created a time series of mouth-looking time. As adjacent time samples exhibit similar differences, we made comparisons using a statistical test based on clustering of adjacent time samples (Maris & Oostenveld, 2007).

Generalized linear mixed model analysis

To further explore whether participants' face-looking patterns could predict their responses, we performed a logistic mixed model analysis using R 3.3.3 (R Core Team, 2017). We used participants' looking time on each AOI to predict their responses (i.e., McGurk or non-McGurk percept) in each trial. That is, we took participants' looking time on the eyes, the mouth, the nose, and the other area (i.e., the area on the face except for the eyes, nose, and mouth) as fixed effects and participants as a random effect to predict participants' responses. The model was as follows:

$$\text{Response} = \beta_0 + \beta_1 * \text{eyes} + \beta_2 * \text{mouth} + \beta_3 * \text{nose} + \beta_4 * \text{other} + \text{participants}^a$$

[a] Each of the AOI names (eyes, mouth, nose, other) in the model denoted the total looking time on that AOI.

RESULTS

Weaker McG rk effec in he a i ic gro p han he TD gro p

We used Mann–Whitney U tests to compare the group differences in the accuracies in congruent trials of the two conditions. The results showed no significant group difference in the accuracies in the open-eyes condition, $U = 315.00$, $Z = 0.52$, $p = 0.611$, $r = 0.07$ ($M_{\text{autistic}} = 0.98$, $SD_{\text{autistic}} = 0.04$; $M_{\text{TD}} = 0.98$, $SD_{\text{TD}} = 0.03$), or in the closed-eyes condition, $U = 353.00$, $Z = -0.54$, $p = 0.604$, $r = -0.07$ ($M_{\text{autistic}} = 0.98$, $SD_{\text{autistic}} = 0.05$; $M_{\text{TD}} = 0.996$, $SD_{\text{TD}} = 0.01$).

We conducted the same tests to examine the group differences of the responses in the incongruent trials of the two conditions. In the open-eyes condition, the results showed that compared with the TD group, the autistic group made fewer "da" responses (McGurk responses), $U = 476.00$, $Z = -2.60$, $p = 0.036$, $r = -0.36$, more "ba" responses, $U = 214.50$, $Z = 2.28$, $p = 0.046$, $r = 0.32$, similar amounts of "ga," $U = 323.50$, $Z = 0.37$, $p = 0.71$, $r = 0.05$, and similar amounts of other responses, $U = 299.00$, $Z = 1.40$, $p = 0.35$, $r = 0.19$ (Figure 3(a)). In the closed-eyes condition, the results showed that the two groups made similar amounts of "ba," $U = 275.50$, $Z = 1.18$, $p = 0.40$, $r = 0.16$, "da" $U = 401.00$, $Z = -1.23$, $p = 0.55$, $r = -0.17$, "ga," $U = 351.50$, $Z = 0.34$, $p = 0.92$, $r = 0.05$, and other responses, $U = 312.00$, $Z = 1.43$, $p = 0.77$, $r = 0.20$ (all ps were corrected by FDR correction; Figure 3(b)).

To further test the group and condition differences of the McGurk effect ("da" response), we conducted a two-way repeated measures permutation ANOVA with Group as the between-subject factor and Condition (open-eyes vs. closed-eyes) as the within-subject factor using the *R* package "permuco" default method (Frossard & Renaud, 2019). The *R* package "permuco" can be used to conduct permutation tests for (repeated measures) ANOVA/ANCOVA, regression and comparison of signals. The results showed a marginally significant main effect of Group, $F(1, 50) = 3.83$, permutation $p = 0.051$, $\eta_p^2 = 0.07$, a significant main effect of Condition, $F(1, 50) = 4.39$, permutation $p = 0.044$, $\eta_p^2 = 0.08$, and a significant Group × Condition interaction, $F(1, 50) = 4.39$, permutation $p = 0.040$, $\eta_p^2 = 0.08$. Further comparison using the Wilcoxon signed rank test showed that the TD group showed a greater McGurk effect in the open-eyes condition than in the closed-eyes condition, $z = 2.36$, $p = 0.018$, $r = 0.44$, and that the autistic group showed a similar McGurk effect in these two conditions, $z = -0.07$, $p = 0.94$, $r = 0.027$ (Figure 4).

In sum, we found that the autistic group showed a weaker McGurk effect than the TD group in the open-eyes condition, while the two groups showed similar McGurk effect in the closed-eyes condition. The autistic

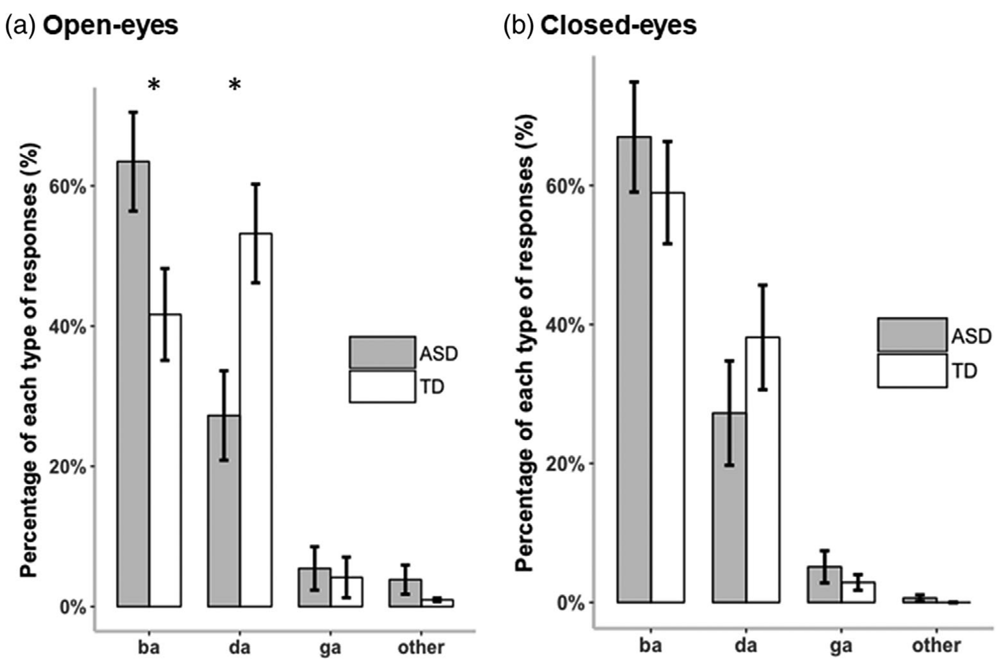

FIGURE 3 Percentages of the two groups' responses in the open-eyes condition (a) and the closed-eyes condition (b). The horizontal axes denote the four types of responses made by participants. The vertical axes denote the percentage of each type of responses. Error bars represent standard errors of the mean (SEMs). *$p < 0.05$

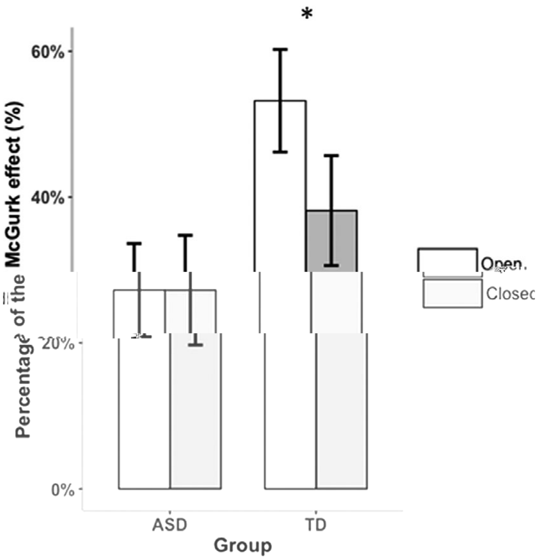

FIGURE 4 Percentages of the McGurk effect (percent of "da" response) in the open-and closed-eyes conditions in the two groups. Error bars were SEMs. $^{*}p < 0.05$

group showed a similar McGurk effect in the two conditions, while the TD group showed a weaker McGurk effect in the closed-eyes condition than in the open-eyes condition.

Temporal-course analysis for the mouth-looking time

We analyzed the proportional looking time spent on each AOI and compared it between groups and conditions (Figure S1). We also conducted a repeated measures ANOVA to examine the effect of condition and group on mouth-looking time with Group as the between-subject factor and Condition (open-eyes vs. closed-eyes) as the within-subject factor. Results showed significant main effects of Group, $F(1, 50) = 6.70$, $p = 0.013$, $\eta_p^2 = 0.12$, and Condition, $F(1, 50) = 4.42$, $p = 0.040$, $\eta_p^2 = 0.08$, but the Group × Condition interaction did not reach significance, $F(1, 50) = 1.13$, $p = 0.29$, $\eta_p^2 = 0.02$. Moreover, we compared the mouth-looking time in the two groups between the two conditions. The results revealed that autistic children showed longer mouth-looking time in the closed-eyes condition ($M = 0.66$, $SD = 0.29$) than in the open-eyes condition ($M = 0.56$, $SD = 0.26$), $t(26) = 2.15$, $p = 0.041$, Cohen's $d = 0.35$, and that TD children showed similar mouth-looking time in the open-eyes condition ($M = 0.83$, $SD = 0.40$) and the closed-eyes condition ($M = 0.84$, $SD = 0.39$), $t(26) = 0.28$, $p = 0.78$, Cohen's $d = 0.036$. That is, closed-eye manipulation affected the mouth-looking time in autistic children but not in TD children when processing audiovisual incongruent trials. In addition, we examined the effect of group and condition on the time looking away from the face, and only found a significant main effect of Group, $F(1, 50) = 5.64$, $p = 0.02$, $\eta_p^2 = 0.10$ (see Supporting Information for detailed results). We further tested the effect of group and condition on the ratio of mouth: eyes looking time. But no significant result was found (see Supporting Information for detailed results).

Viewing the mouth has been shown to be critical for evoking audiovisual speech integration (McGurk effect; Gurler et al., 2015). As each stimulus lasted as long as approximately 2.10 s in our study, participants' mouth-looking behavior possibly changed over time. Thus, we further explored the temporal course of mouth-looking time in the incongruent trials and compared it between groups. We found that in the open-eyes condition, the autistic group's mouth-looking time was shorter than that of the TD group during almost the whole latter half period of the stimulation (0.88–1.26 s, $p = 0.002$, and 1.3–1.78 s, $p = 0.002$, see Figure 5(a)). During this period, the speaker first opened her mouth to articulate the phonemes and then closed her mouth. In the closed-eyes condition, no difference was found between groups, $p > 0.05$ (see Figure 5(b)).

Additionally, we examined whether the temporal course of the mouth-looking time differed for the McGurk and non-McGurk trials within each group. Autistic children spent more time viewing the mouth in the McGurk trials than in the non-McGurk trials from 0.82 s to 1.03 s in the open-eyes condition, $p = 0.002$ (Figure 6(a)), and from 0.37 s to 1.53 s in the closed-eyes condition, $p = 0.002$ (Figure 6(c)). Within these periods, the speaker opened her mouth to articulate phonemes. The TD group showed no difference between the McGurk and non-McGurk trials in the open- or closed-eyes conditions, $ps > 0.05$ (Figure 6(b),(d)). Results for the temporal course analysis of looking time for other AOIs were shown in Figures S2–S5.

In sum, the autistic group showed a weaker McGurk effect, and they spent less time viewing the mouth than the TD group when the speaker articulated the phoneme in the open-eyes condition. The two groups showed similar McGurk effect and spent similar amounts of time viewing the mouth in the closed-eyes condition. The autistic group spent less time viewing the mouth in the non-McGurk trials than they did in the McGurk trials when the speaker articulated the phonemes in both conditions. These findings indicated that the weaker McGurk effect in the autistic group was probably related to their reduced mouth-looking time in both conditions.

Face-viewing pattern predicts McGurk response

We conducted a generalized linear mixed model analysis to further explore whether participants' looking patterns (e.g., mouth-looking time, eye-looking time) could predict their McGurk responses. In the open-eyes condition,

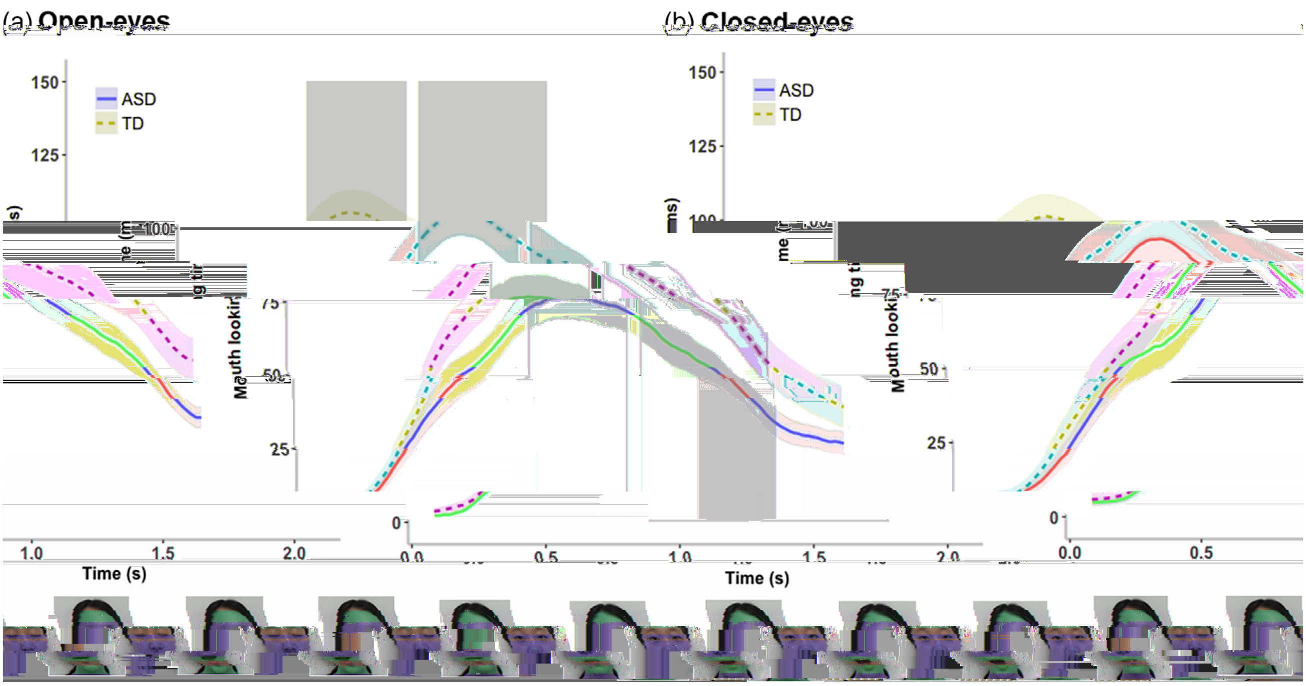

FIGURE 5 Mouth-looking time in the ASD and TD groups for the incongruent trials over time in the open-eyes condition (a) and the closed-eyes condition (b). The vertical axes denote the mouth-looking time within an epoch of 1/6 s, the horizontal axes denote the timeline of the videos, and the images below the horizontal axis show the speaker's mouth movements at 0, 0.5, 1, 1.5, and 2.0 s. Shaded areas indicate SEMs. Gray areas illustrate the time epochs during which the mouth-looking time was significantly different between the two groups at $p < 0.05$

as shown in Table 2, the autistic group's mouth-looking time could significantly predict their McGurk response, $\beta_2 = 0.82$, $SEM = 0.42$, $p = 0.049$, and the autistic group's nose-looking time was a marginally significant factor in predicting their McGurk response, $\beta_3 = 0.88$, $SEM = 0.46$, $p = 0.054$; the TD group's eye-looking time could significantly predict their McGurk response, $\beta_1 = 1.47$, $SEM = 0.57$, $p = 0.0098$ (see Table 2 for detailed information). In the closed-eyes condition, the autistic group's mouth-looking time could significantly predict their McGurk response, $\beta_2 = 1.99$, $SEM = 0.57$, $p = 0.0005$, and the autistic group's looking time for the other area could significantly predict their McGurk response, $\beta_4 = 2.38$, $SEM = 0.86$, $p = 0.005$; none of the TD group's looking time on three AOIs could predict their McGurk response, ps > 0.05. That is, the McGurk effect of the autistic group could be predicted by their mouth-looking time in both conditions, and the McGurk effect of the TD group could be predicted by their eye-looking time in the open-eyes condition but could not be predicted by their looking time on any of the four AOIs in the closed-eyes condition.

DISCUSSION

In the present study, by employing the McGurk effect, we explored audiovisual speech integration, face viewing patterns, and the association between them in autistic children and TD children. We found that, compared with TD children, autistic children showed weaker audiovisual speech integration (McGurk effect) and atypical face viewing patterns and that the mouth-looking time in autistic children could predict their audiovisual speech integration in both conditions. For TD children, the eye-looking time could predict their audiovisual speech integration in the open-eyes condition. We also observed that audiovisual speech integration in autistic children was similar in the open- and the closed-eyes conditions, but audiovisual speech integration in TD children was weaker in the closed-eyes condition than in the open-eyes condition.

Our key findings were as follows. First, consistent with our hypothesis, we observed that autistic children showed weaker audiovisual speech integration (McGurk effect), which is consistent with most previous studies (Bebko et al., 2014; de Gelder et al., 1991; Irwin et al., 2011; Mongillo et al., 2008; Stevenson et al., 2014). The weaker audiovisual speech integration in autistic children could be explained by the weak central coherence theory, which emphasizes inferiority in holistic processing in autistic children (Happé & Frith, 2006). In the present study, autistic children showed difficulty processing audiovisual speech holistically; that is, they made less fused McGurk perceptions.

Second, temporal course analysis revealed that the autistic group showed less mouth-looking time than the TD group in the open-eyes condition. When we

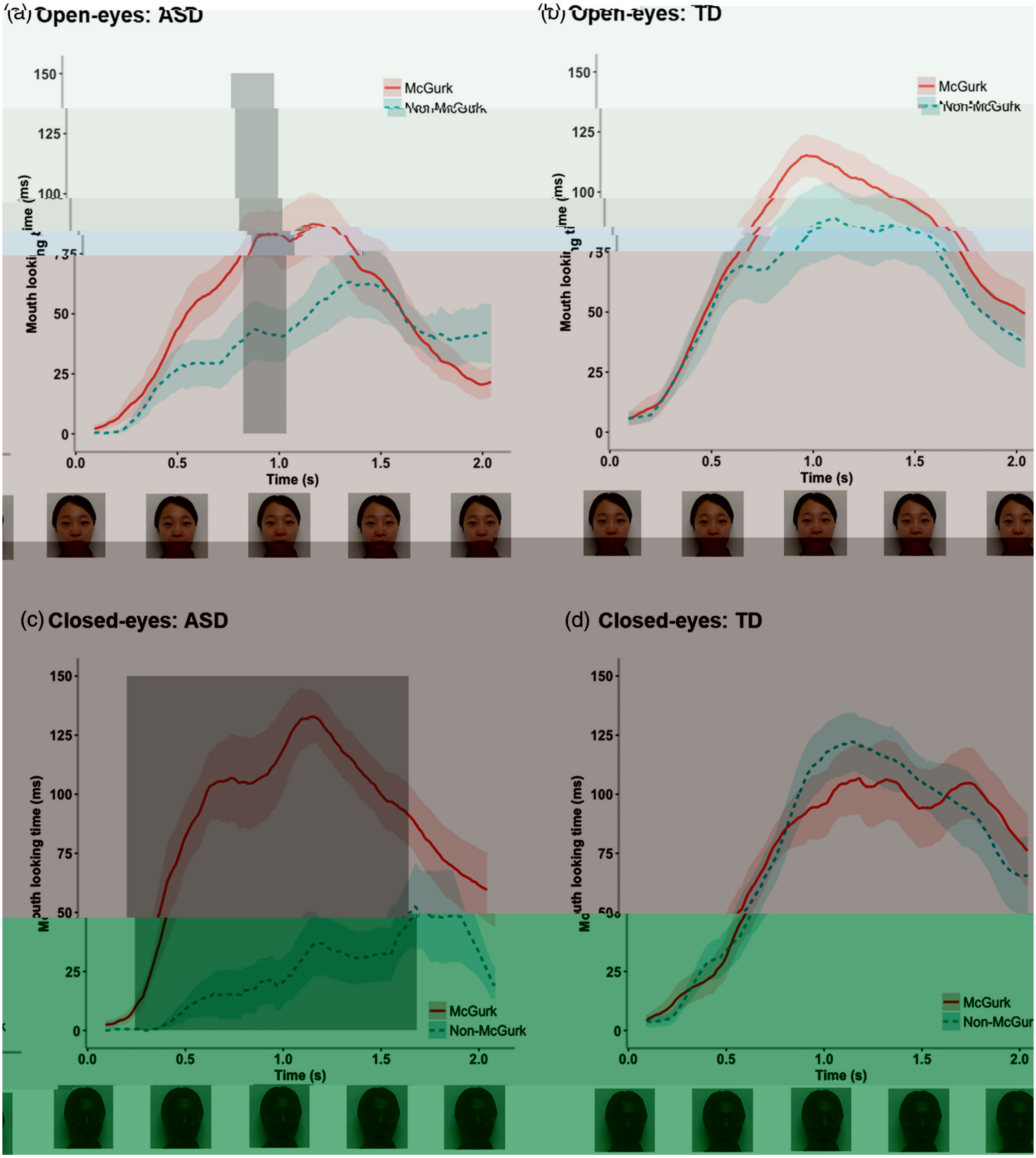

FIGURE 6 Mouth-looking times between the McGurk and non-McGurk trials over time within the ASD and TD groups in the open-eyes condition (a, b) and the closed-eyes condition (c, d). The vertical axes denote the mouth-looking time within an epoch of 1/6 s, the horizontal axes denote the timeline of the videos, and the images below the horizontal axis show the speaker's mouth movement at 0, 0.5, 1, 1.5, and 2.0 s. Shaded areas indicate SEMs. The gray area illustrates the time epochs at which the mouth-looking time was significantly different between the two kinds of trials at $p < 0.05$

examined the effect of condition and group on mouth-looking time, the significant effect of Group also indicated less mouth-looking time in the autistic group than in the TD group, though no significant interaction was found. Our finding that the autistic group viewed the mouth less confirms our hypothesis and conforms to the findings of most previous studies (Chawarska & Shic, 2009; de Wit et al., 2008; Nakano et al., 2010). It

TABLE 2 Results of the generalized linear mixed model analysis. β_0 is the intercept, and β_1, β_2, β_3, and β_4 are the coefficients of the looking time on the eyes, mouth, and nose, and other area respectively, in the model

Condi ion	Gro p	LL[a]	De iance	β_0 (*SEM*[b])	β_1 (*SEM*)	β_2 (*SEM*)	β_3 (*SEM*)	β_4 (*SEM*)
Open-eyes	ASD	−131.1	262.1	−2.21** (0.76)	0.72 (0.46)	0.82* (0.42)	0.88† (0.46)	0.51 (0.72)
	TD	−155.8	311.6	−0.23 (0.53)	1.47** (0.57)	0.49 (0.30)	−0.30 (0.71)	−0.10 (0.56)
Closed-eyes	ASD	−87.4	174.8	−5.43* (2.28)	−0.05 (0.97)	1.99*** (0.57)	−1.50 (1.12)	2.38** (0.86)
	TD	−135.0	270.0	−1.20 (0.81)	−0.08 (0.44)	0.43 (0.38)	0.12 (0.55)	0.76 (0.67)

†$p < 0.08$.
*$p < 0.05$; **$p < 0.01$; ***$p < 0.001$.
[a]Log likelihood.
[b]Standard error of the mean.

has been found that autistic children viewed the speaker's mouth less when viewing videos of talking people (Nakano et al., 2010). Moreover, less gaze to the mouth in young children has been found to predict their later lower levels of expressive language (Young et al., 2009). Thus, autistic children's shorter mouth-looking time might help explain their delayed expressive language ability and hence their social communication deficits. But it needs to be confirmed in future studies.

Third, we further confirmed our hypothesis by discovering that the less mouth-looking time in the autistic group could hypothetically predict their audiovisual speech integration deficits. This was further proven by our finding that the autistic group viewed the mouth less in the non-McGurk trials than in the McGurk trials in both conditions. It is noteworthy that the reduced mouth-looking time in the autistic group was in the context of their reduced face-looking time or their increased time looking away from the face compared with TD children. These findings indicate that audiovisual speech integration deficits in autistic children could be partially accounted for by their atypical visual attention allocation hypothetically. In the autistic group, we also found that nose-looking time in the open-eyes condition or time looking at the other area in the closed-eyes condition could also predict their audiovisual speech integration. This might be due to the fact that the autistic group could process the speaker's mouth movements through their peripheral vision when they viewing these two areas which were close to the mouth area (Strasburger et al., 2011). In the same vein, similar effects of visual attention on audiovisual speech integration have been found in TD individuals, who showed weaker audiovisual speech integration as their visual attention was disturbed by a visual distractor (Tiippana et al., 2004). Furthermore, the strength of audiovisual speech integration in TD adults has been found positively correlated with their duration of visual attention to the mouth (Gurler et al., 2015). The combined evidence could be explained by a hypothesis of "the interplay between attention and multisensory integration," in which attention affects multisensory integration (e.g., audiovisual integration) and, in turn, multisensory integration affects attention (Talsma et al., 2010). In the autistic group, face viewing pattern could predict audiovisual speech integration, indicating that attention affects multisensory integration hypothetically. This finding disclosed that the atypical face viewing pattern was one of the underlying mechanisms of audiovisual speech integration deficits in autistic children. In this way, we could try to improve audiovisual speech integration in autistic children by directing their visual attention to the speaker's mouth (i.e., increasing their mouth-looking time). Our findings might also provide insights for explaining the social and communicative deficits in autistic children. The atypical face viewing patterns in autistic children impair their audiovisual speech integration. The impaired audiovisual speech integration might further impede their language ability development in their early life by affecting speech comprehension and language processing during social interactions (Gervain & Mehler, 2010; Werker & Gervain, 2013). Consequently, the delayed language ability in autistic children might impair their social communication. It needs to directly and systematically explore the relationship among face viewing patterns, audiovisual speech integration, and social communication in autistic children in future studies.

In addition, for the TD group, we found that audiovisual speech integration could be hypothetically predicted by eye-looking time in the open-eyes condition. The hypothetical prediction of eye-looking time for audiovisual speech integration in the TD group could be accounted for by the fact that the communicative information conveyed by the speaker's eyes could facilitate speech perception including audiovisual speech integration (Davidhizar, 1992; Holler et al., 2014). A previous study also found that mouth-looking time was linked with audiovisual speech integration in TD adults (Gurler et al., 2015). Combining these findings, we speculated that visual attention to the core facial features is essential for audiovisual speech integration in TD individuals. Future studies could further explore the association between face viewing patterns and audiovisual speech integration in different ages and populations.

Fourth, the audiovisual speech integration in the TD and autistic groups was impacted differently by the closed-eyes condition. For the TD group, the audiovisual speech integration was weakened in the closed-eyes condition compared with the open-eyes condition, though the TD group demonstrated similar mouth-looking time in the two conditions. This finding might be explained by two reasons. First, the talking face with closed-eyes is a rather unusual face that might evoke an extra cognitive processing and trigger extra visual scanning, which might weaken the audiovisual speech integration in TD children though their mouth-looking time kept stable. Second, the closed-eyes setting eliminated eyes' information which is crucial for audiovisual speech integration in TD children. Eyes' information could facilitate the audiovisual speech integration in TD children as our results showed that eyes-looking time predicted the audiovisual speech integration in TD children in the open-eyes condition. This is also consistent with previous findings that eyes information could facilitate speech perception (Holler et al., 2014). Unlike the TD group, the audiovisual speech integration in the autistic group was similar in the two conditions. This might be because that, in the open-eyes condition, the autistic group could not effectively process the eyes' information, which is crucial for audiovisual speech integration (Baron-Cohen et al., 1997). This might be another possible mechanism underlying audiovisual speech integration deficits in autistic children.

The present study also has some limitations. For example, participants were all boys, and we could not generalize our conclusions to girls. In future studies, we could explore whether atypical face viewing patterns in autistic girls could predict their audiovisual speech integration deficits. We could also explore whether the closed-eyes condition could weaken the audiovisual speech integration in TD girls.

In summary, we found that the atypical face viewing patterns in autistic children could predict their audiovisual speech integration deficits. This finding indicates that audiovisual speech integration in autistic children is affected by their face viewing patterns, especially the looking time on the mouth. It unveiled that one of the underlying mechanisms of audiovisual speech integration deficits in autistic children was their atypical face viewing patterns. These findings indicate an explanation for the social communication deficits in autistic children from the perspective of social attention. That is, the social attention deficits in autistic children impair their audiovisual speech integration, which further impairs their language ability and hence their social communication. Our findings also provide supporting evidence for "the interplay between attention and multisensory integration" (Talsma et al., 2010) and extend its application to autistic children. Our findings further suggest that we could improve audiovisual speech integration in autistic children by manipulating their social attention (i.e., directing their attention to the speaker's mouth).

ACKNOWLEDGMENTS

The authors are grateful to Zipeng Ma, Fuli Liu, Yanhong Wu, Yixiao Hu, Yinan Lv, and the staff in Qingdao Elim School, for their generous assistance in completing the study.

CONFLICT OF INTEREST

The authors declare no potential conflict of interest.

ETHICS STATEMENT

This study was approved by the research ethics committee at Peaking University.

ORCID

Tianbi Li https://orcid.org/0000-0002-9343-1461
Li Yi https://orcid.org/0000-0003-4117-8828

REFERENCES

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). American Psychiatric Press.

Auyeung, B., Baron-Cohen, S., Wheelwright, S., & Allison, C. (2008). The autism spectrum quotient: Children's version (AQ-child). *Journal of Autism and Developmental Disorders*, *38*, 1230–1240. https://doi.org/10.1007/s10803-007-0504-z

Baron-cohen, S., Jolliffe, T., Mortimore, C., & Robertson, M. (1997). Another advanced test of theory of mind: Evidence from very high functioning adults with autism or Asperger syndrome. *Journal of Child Psychology and Psychiatry*, *38*, 813–822. https://doi.org/10.1111/j.1469-7610.1997.tb01599.x

Baum, S. H., Stevenson, R. A., & Wallace, M. T. (2015). Behavioral, perceptual, and neural alterations in sensory and multisensory function in autism spectrum disorder. *Progress in Neurobiology*, *134*, 140–160. https://doi.org/10.1016/j.pneurobio.2015.09.007

Bebko, J. M., Schroeder, J. H., & Weiss, J. A. (2014). The McGurk effect in children with autism and Asperger syndrome. *Autism Research*, *7*, 50–59. https://doi.org/10.1002/aur.1343

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision*, *10*, 433–436. https://doi.org/10.1163/156856897X00357

Chawarska, K., & Shic, F. (2009). Looking but not seeing: Atypical visual scanning and recognition of faces in 2 and 4-year-old children with autism spectrum disorder. *Journal of Autism Developmental Disorder*, *39*, 1663–1672. https://doi.org/10.1007/s10803-009-0803-7

Dankner, Y., Shalev, L., Carrasco, M., & Yuval-Greenberg, S. (2017). Prestimulus inhibition of saccades in adults with and without attention deficit/hyperactivity disorder as an index of temporal expectations. *Psychological Science*, *28*, 835–850. https://doi.org/10.1177/0956797617694863

Davidhizar, R. (1992). Interpersonal communication: A review of eye contact. *Infection Control and Hospital Epidemiology*, *13*(4), 222–225. https://doi.org/10.1086/646513

de Gelder, B., Vroomen, J., & van der Heide, L. (1991). Face recognition and lip-reading in autism. *European Journal of Cognitive Psychology*, *3*, 69–86. https://doi.org/10.1080/09541449108406220

de Wit, T. C. J., Falck-Ytter, T., & von Hofsten, C. (2008). Young children with autism spectrum disorder look differently at positive versus negative emotional faces. *Research in Autism Spectrum Disorders*, *2*, 651–659. https://doi.org/10.1016/j.rasd.2008.01.004

Feldman, J. I., Dunham, K., Cassidy, M., Wallace, M. T., Liu, Y., & Woynaroski, T. G. (2018). Audiovisual multisensory integration in individuals with autism spectrum disorder: A systematic review and meta-analysis. *Neuroscience Biobehavior Reviews*, *95*, 220–234. https://doi.org/10.1016/j.neubiorev.2018.09.020

Fort, M., Kandel, S., Chipot, J., Savariaux, C., Granjon, L., & Spinelli, E. (2013). Seeing the initial articulatory gestures of a word triggers lexical access. *Language and Cognitive Processes*, *28*, 1207–1223. https://doi.org/10.1080/01690965.2012.701758

Frossard, J. & Renaud, O. (2019). permuco: Permutation tests for regression, (repeated measures) ANOVA/ANCOVA and comparison of signals. R package version 1.1.0. https://CRAN.R-project.org/package=permuco

Gervain, J., & Mehler, J. (2010). Speech perception and language acquisition in the first year of life. *Annual Review of Psychology*, *61*(1), 191–218. https://doi.org/10.1146/annurev.psych.093008.100408

Gurler, D., Doyle, N., Walker, E., Magnotti, J., & Beauchamp, M. (2015). A link between individual differences in multisensory speech perception and eye movements. *Attention, Perception, & Psychophysics*, *77*, 1333–1341. https://doi.org/10.3758/s13414-014-0821-1

Happé, F., & Frith, U. (2006). The weak coherence account: Detail-focused cognitive style in autism spectrum disorders. *Journal of Autism and Developmental Disorders*, *36*(1), 5–25. https://doi.org/10.1007/s10803-005-0039-0

Holler, J., Schubotz, L., Kelly, S., Hagoort, P., Schuetze, M., & Özyürek, A. (2014). Social eye gaze modulates processing of speech and co-speech gesture. *Cognition*, *133*(3), 692–697. https://doi.org/10.1016/j.cognition.2014.08.008

Iarocci, G., Rombough, A., Yager, J., Weeks, D. J., & Chua, R. (2010). Visual influences on speech perception in children with autism. *Autism*, *14*, 305–320. https://doi.org/10.1177/1362361309353615

Irwin, J. R., Tornatore, L. A., Brancazio, L., & Whalen, D. H. (2011). Can children with autism spectrum disorders "hear" a speaking face? *Child Development*, *82*, 1397–1403. https://doi.org/10.1111/j.1467-8624.2011.01619.x

Jack, R. E., & Schyns, P. G. (2015). The human face as a dynamic tool for social communication. *Current Biology*, *25*, 621–634. https://doi.org/10.1016/j.cub.2015.05.052

Kleiner, M., Brainard, D., Pelli, D., Ingling, A., Murray, R., & Broussard, C. (2007). What's new in psychtoolbox-3? *Perception*, *36*, 1–16.

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods*, *164*, 177–190. https://doi.org/10.1016/j.jneumeth.2007.03.024

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, *264*, 746–748. https://doi.org/10.1023/A:1005592401947

Mongillo, E. A., Irwin, J. R., Whalen, D. H., Klaiman, C., Carter, A. S., & Schultz, R. T. (2008). Audiovisual processing in children with and without autism spectrum disorders. *Journal of Autism & Developmental Disorders*, *38*, 1349–1358. https://doi.org/10.1007/s10803-007-0521-y

Nakano, T., Tanaka, K., Endo, Y., Yamane, Y., Yamamoto, T., Nakano, Y., Ohta, H., Kato, N., & Kitazawa, S. (2010). Atypical gaze patterns in children and adults with autism spectrum disorders dissociated from developmental changes in gaze behaviour. *Proceedings of the Royal Society B*, *277*, 2935–2943. https://doi.org/10.1098/rspb.2010.0587

Olsen, A. (2012). *The Tobii I-VT fixation filter: Algorithm Description.* Tobii I-VT Fixation Filter-white paper *[white paper]*. Stockholm: Tobii Technology. http://www.tobii.com/no/eye-tracking-research/global/library/white-papers/the-tobii-i-vt-fixation-filter/.

Pascalis, O., Loevenbruck, H., Quinn, P. C., Kandel, S., Tanaka, J. W., & Lee, K. (2014). On the links among face processing, language processing, and narrowing during development. *Child Development Perspectives*, *8*(2), 65–70. https://doi.org/10.1111/cdep.12064

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, *10*, 437–442. https://doi.org/10.1163/156856897x00366

Pelphrey, K. A., Sasson, N. J., Reznick, J. S., Paul, G., Goldman, B. D., & Piven, J. (2002). Visual scanning of faces in autism. *Journal of Autism & Developmental Disorders*, *32*, 249–261. https://doi.org/10.1016/j.tics.2012.08.009

R Core Team. (2017). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. http://www.R-project.org/

Stevenson, R. A., Siemann, J. K., Schneider, B. C., Eberly, H. E., Woynaroski, T. G., Camarata, S. M., & Wallace, M. T. (2014). Multisensory temporal integration in autism spectrum disorders. *Journal of Neuroscience*, *34*, 691–697. https://doi.org/10.1523/JNEUROSCI.3615-13.2014

Strasburger, H., Rentschler, I., & Jüttner, M. (2011). Peripheral vision and pattern recognition: A review. *Journal of Vision*, *11*(5), 13–13.

Sumby, W. H., & Pollack, I. (1954). Visual contribution to speech intelligibility in noise. *The Journal of the Acoustical Society of America*, *26*, 212–215. https://www.sci-hub.shop/, https://doi.org/10.1121/1.1907309

Talsma, D., Senkowski, D., Soto-Faraco, S., & Woldorff, M. G. (2010). The multifaceted interplay between attention and multisensory integration. *Trends in Cognitive Sciences*, *14*, 400–410. https://doi.org/10.1016/j.tics.2010.06.008

Tanaka, J. W., & Sung, A. (2016). The "eye avoidance" hypothesis of autism face processing. *Journal of Autism & Developmental Disorders*, *46*, 1538–1552. https://doi.org/10.1007/s10803-013-1976-7

Tiippana, K., Andersen, T. S., & Sams, M. (2004). Visual attention modulates audiovisual speech perception. *European Journal of Cognitive Psychology*, *16*(3), 457–472. https://doi.org/10.1080/09541440340000268

Wechsler, D. (2014). *Wechsler preschool and primary scale of intelligence—Fourth CN edition (WPPSI-IV CN)*. (Y. Li & Z. J, Trans. Eds.). China: King-May Company.

Werker, J. F., & Gervain, J. (2013). Speech perception in infancy: A foundation for language acquisition. In P. D. Zelazo (Ed.), *The Oxford handbook of developmental psychology* (pp. 909–925). Oxford Handbooks Online. https://doi.org/10.1093/oxfordhb/9780199958450.013.0031

Woynaroski, T. G., Kwakye, L. D., Foss-Feig, J. H., Stevenson, R. A., Stone, W. L., & Wallace, M. T. (2013). Multisensory speech perception in children with autism spectrum disorders. *Journal of Autism and Developmental Disorders*, *43*(12), 2891–2902. https://doi.org/10.1007/s10803-013-1836-5

Young, G. S., Merin, N., Rogers, S. J., & Ozonoff, S. (2009). Gaze behavior and affect at 6 months: Predicting clinical outcomes and language development in typically developing infants and infants at risk for autism. *Developmental Science*, *12*(5), 798–814. https://doi.org/10.1111/j.1467-7687.2009.00833.x

Zhang, J., Meng, Y., He, J., Xiang, Y., Wu, C., Wang, S., & Yuan, Z. (2019). McGurk effect by individuals with autism spectrum disorder and typically developing controls: A systematic review and meta-analysis. *Journal of Autism and Developmental Disorders*, *49*, 34–43. https://doi.org/10.1007/s10803-018-3680-0

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

Ho o ci e hi ar icle: Feng, S., Lu, H., Wang, Q., Li, T., Fang, J., Chen, L., & Yi, L. (2021). Face-viewing patterns predict audiovisual speech integration in autistic children. *Autism Research*, 1–11. https://doi.org/10.1002/aur.2598