Received: 14 January 2022 | Revised: 30 September 2022 | Accepted: 10 November 2022

DOI: 10.1111/desc.13348

Developmental Science WILEY

RESEARCH ARTICLE

Increasing audiovisual speech integration in autism through enhanced attention to mouth

Shu uan Feng[1,2] | Qiandong Wang[3] | Yi iao Hu[2] | Hao ang Lu[2,4,5] | Tianbi Li[2] | Ci Song[2] | Jing Fang[6] | Lihan Chen[2,7] |

Highlights

- The present study examined whether audiovisual speech integration in the McGurk task in AC could be increased by increasing their attention to the speaker's mouth.
- Blurring the speaker's eyes increased mouth-looking time and audiovisual speech integration in the McGurk task in AC.
- Cuing to the speaker's mouth also increased mouth-looking time and audiovisual speech integration in the McGurk task in AC.
- Audiovisual speech integration in the McGurk task in AC could be increased by increasing their attention to the speaker's mouth.

1 | INTRODUCTION

Audiovisual speech integration entails the combination of auditory and visual parts of a speech into a coherent representation (Altieri et al., 2011). Reduced audiovisual integration in McGurk tasks has been reported in autistic children (AC) (Bebko et al., 2014; Iarocci et al., 2010; Irwin et al., 2011; Mongillo et al., 2008; Stevenson et al., 2014). Autism is a neurodevelopmental condition characterized by difficulties in social interactions and communications, as well as restricted and repetitive patterns of behavior (DSM-5; American Psychiatric Association, 2013). The reduced audiovisual speech integration in AC was associated with language or communication difficulties (Feldman et al., 2018).

Audiovisual speech integration has been measured by susceptibility to the McGurk effect, which occurs when the auditory part of a phoneme is dubbed onto the mouth movements of another (visually presented) phoneme, leading to a fused perception of a new phoneme (McGurk & MacDonald, 1976). For example, when the auditory phoneme "ba" was dubbed onto the visual mouth movements of "ga," people often reported a fused perception of "da" (McGurk & MacDonald, 1976). Using the McGurk effect paradigm, a series of studies investigated audiovisual speech integration in AC (Bebko et al., 2014; Iarocci et al., 2010; Irwin et al., 2011; Mongillo et al., 2008; Stevenson et al., 2014; Woynaroski et al., 2013). A recent meta-analysis summarized that AC have less audiovisual speech integration (i.e., less McGurk effect; Zhang et al., 2019).

Audiovisual speech integration, a typical form of multisensory integration, has been linked to and can be modulated by attention (e.g., Talsma et al., 2010). Some studies have explored how attention affects audiovisual speech integration by employing the McGurk effect (Alsius et al., 2005, 2007; Tiippana et al., 2004). The McGurk effect was weakened by dual tasks that divided participants' attention, in which observers were distracted by task-irrelevant discrimination of auditory (Alsius et al., 2005), visual (Alsius et al., 2005; Tiippana et al., 2004), or tactile stimuli (Alsius et al., 2007). Other studies have further revealed the possible relationship between the McGurk effect and participants' visual attention (e.g., Feng et al., 2021; Gurler et al., 2015). It has been further proved that the strength of the McGurk effect was correlated with individuals' attention to the speaker's core facial features, such as their attention to the speaker's mouth (i.e., mouth-looking time; Gurler et al., w6.11m[(7T-222GS3g0.0004Tc0TLTc7.99701Specifegr)ly,7.997017.97[illegible]re

TABLE 1 Participants' characteristics of the autistic and nonautistic groups

			N	Male/female	Mean age in ears (SD)	IQ (WPPSI-IV)
Experiment 1	Autistic		30	30/0	5.65 (0.77)	111.00 (11.82)
	Nonautistic		30	30/0	5.73 (0.51)	111.30 (10.73)
	t (*p*)	Autistic vs. Nonautistic	N/A	N/A	−0.45 (0.65)	−0.10 (0.92)
Experiment 2	Autistic		40	40/0	5.55 (0.68)	113.00 (10.56)
	Nonautistic		42	42/0	5.71 (0.59)	109.29 (10.37)
	t (*p*)	Autistic vs. Nonautistic			−1.14 (.26)	1.61 (0.11)

Note. IQ was measured using the Chinese version of the Wechsler Preschool and Primary Scale of Intelligence-Fourth Edition (WPPSI-IV). All *p*s > 0.05.

speakers was blurred. We hypothesized that (a) blurring the speaker's eyes could weaken the attractiveness of eye regions, and thus could decrease the eyes-looking time and increase the mouth-looking time; and (b) this increased mouth-looking time would increase audiovisual speech integration in the McGurk task in AC. Experiment 2 included three conditions: cue-to-mouth, cue-to-eyes, and free-viewing. We directed children's attention to the mouth in the cue-to-mouth condition and to the eyes in the cue-to-eyes condition and allowed them to view the screen freely in the free-viewing condition. We hypothesized that, compared with the free-viewing condition, (a) the cue-to-mouth condition would increase the mouth-looking time and the cue-to-eyes condition would increase the eyes-looking time in AC; and (b) the increased mouth-looking time in the cue-to-mouth condition would increase the audiovisual speech integration in the McGurk task in AC, but the increased eyes-looking time in the cue-to-eyes condition would not change the audiovisual speech integration in the McGurk task in AC because of their difficulty in processing the eye information (Baron-Cohen et al., 1997).

2 | EXPERIMENT 1

In this experiment, we set a clear-eyes condition and a blurred-eyes condition to explore whether blurring the speaker's eyes could enhance children's audiovisual speech integration in the McGurk task. In these two conditions, we measured children's audiovisual speech integration in the McGurk task and tracked their eye movements.

2.1 | Method

2.1.1 | Participants

We recruited 30 Mandarin-speaking AC (age range: 4.55-7.84 years; all boys) who were from a specialized school for autism. We also recruited 30 mandarin-speaking nonautistic children (NAC; age range: 4.95-6.79 years; all boys) as the comparison group from a kindergarten as well as an elementary school. All AC were diagnosed in licensed hospitals by professional pediatricians according to the criteria of the DSM-V (American Psychiatric Association, 2013). Autism diagnosis was further confirmed according to the Chinese version of the Autism Spectrum Quotient: Children's Version (AQ-Child; Auyeung et al., 2008). In addition, autism screening in the comparison group was also conducted by employing AQ and all children in the comparison group were below the AQ cut-off score (Auyeung et al., 2008). The two groups were matched in both age and intelligence quotient (IQ), revealed by independent samples *t*-tests (see Table 1 for detailed information). IQ was measured using the Chinese version of the Wechsler Preschool and Primary Scale of Intelligence-Fourth Edition (WPPSI-IV; Wechsler, 2014). The study was approved by the research ethics committee of Peking University. Parents of all children signed a written informed consent form before the experiment.

2.1.2 | Stimuli

We used the McGurk effect paradigm (McGurk & MacDonald, 1976) and used syllables identical to those used by Feng et al. (2021) to measure children's audiovisual speech integration in the present study. The experiment included two conditions: clear- and blurred-eyes. Each condition contained congruent and incongruent stimuli. The congruent stimuli were videos of a female speaker uttering "ba" and "ga." The incongruent stimuli were obtained by dubbing the visual "ga" onto the auditory "ba" ("AbVg": auditory "ba" + visual "ga"), which generally evoked the McGurk percept of "da" (McGurk & MacDonald, 1976). Stimuli in the clear-eyes conditions were the original videos we recorded, and stimuli in the blurred-eyes condition were modified by blurring the speaker's eye region in the original videos. Modifications of the stimuli were accomplished using Adobe Premiere Software Pro CS6.0. In the software, we selected Gaussian blur and set the blur parameter to 75%. As the blur parameter increases, the speaker's eye region becomes more blurred. For the practice session, we also prepared three stimuli: "pa," "ka," and "ApVk" (auditory "pa" + visual "ka"). "ApVk" normally evoked the McGurk percept of "ta." The resolution of the videos was 1280 × 720 pixels, with a frame rate of 25 frames/s. We obtained written consent from the female speaker to use these videos.

2.1.3 | Procedures

Children were seated approximately 60 cm from a 21.5-inch Dell screen (resolution: 1920 × 1080 pixels) in a quiet room.

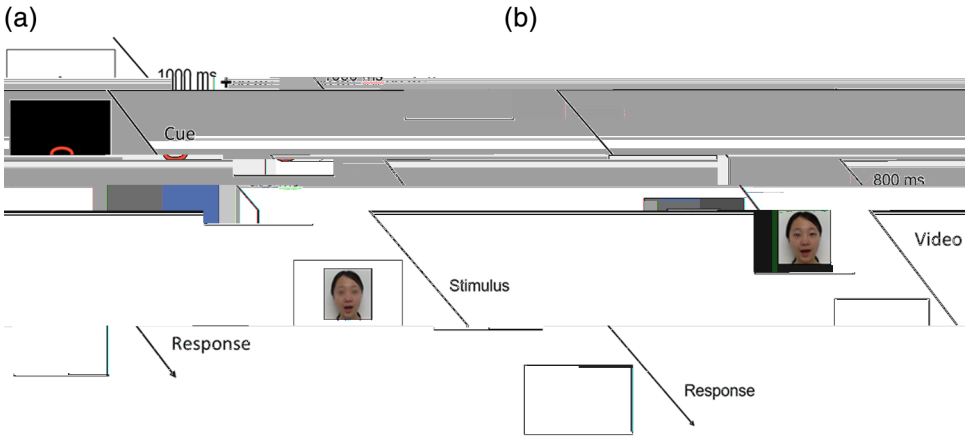

FIGURE 1 Procedures of a sample trial in Experiment 1 (a) and Experiment 2 (b). *Note*. (a) This figure shows the procedure of a trial in Experiment 1. First, a fixation was presented at the center of the screen for 1000 ms. Then, a black screen was displayed for 800 ms. Subsequently, the stimulus was presented. Finally, a black screen was shown, and the children responded. (b) This figure shows the procedure of a sample trial in the cue-to-mouth condition in Experiment 2. First, a black screen with an oval at the position where the speaker's mouth appeared was presented. Then, the stimulus was presented once the children kept fixating on the oval area for 500 ms. Finally, a black screen was displayed until the children responded

The stimuli were displayed at the center of the screen using MATLAB (The MathWorks, Natick, MA) and Psychotoolbox (Brainard, 1997; Kleiner et al., 2007; Pelli, 1997). Sounds were presented through two speakers located on the two sides of the screen. Children were required to perform the McGurk task by reporting what the speaker said, and their eye movements were recorded using a Tobii X 120 eye tracker (sampling rate: 120 Hz).

Children performed a practice session to familiarize themselves with the McGurk task before the formal experiment. At the beginning of the formal experiment, the children's eye movements were calibrated using Tobii's five-point calibration method. The calibration was accepted only when all five points showed a good fit, with error vectors smaller than 0.5 degree of the visual angle. As mentioned above, the experiment consisted of a clear-eyes condition and a blurred-eyes condition. Each condition included four trials of congruent "ba," four trials of congruent "ga," and 12 trials of incongruent "AbVg" (auditory "ba" + visual "ga"). Each trial began with a black fixation at the center of the screen for 1000 ms, and children were asked to look at it. Then, a black screen was displayed for 800 ms. Subsequently, the stimulus was presented. Finally, a black screen was displayed until the children responded. Children's responses were recorded by the experimenter, that is, by pressing "b," "d," and "g" on the keyboard for responses of "ba," "da," and "ga" respectively. For a sample trial procedure, please refer to Figure 1a. The 20 trials in each condition were presented in random order, and the order of the two conditions was counterbalanced among children. Children took rest between the conditions. The experiment lasted for approximately 25 min.

2.1.4 | Data analysis

E e movement data anal sis. We defined five areas of interest (AOIs) for the speaker's face: the whole face, the eyes (left eye and right eye),

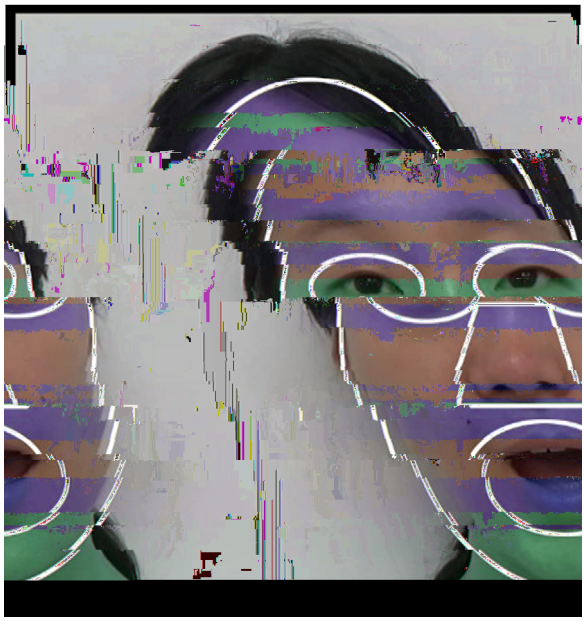

FIGURE 2 Sample AOIs used in the eye movement data analysis. *Note*. This figure shows the five AOIs in the eye-movement data analysis. The five AOIs included the whole face, eyes (left eye and right eye), mouth, nose, and other areas (the area on the face excluding eyes, nose, and mouth)

the mouth, the nose, and the other area (the area on the face excluding eyes, nose, and mouth; see Figure 2). We extracted fixations from the raw gaze data, as specified by Tobii (I-VT fixation filter; Olsen, 2012). In particular, the minimum fixation duration was set at 60 ms within a velocity of 30 deg/s. Then, we obtained the fixation data, which included the onset, the offset, and the position (x-coordinate, y-coordinate in pixels) of each fixation. For each trial, we extracted the fixation data during the time the video was displayed on the screen

(i.e., from the time point that the video appeared on the screen to the time point that the video disappeared on the screen) and calculated the duration of each fixation by using the offset to minus the onset. After that, we selected fixations within each AOI and summed their durations separately, obtaining the total looking time on each AOI. Finally, we calculated the average total looking time on each AOI for each participant and for each group. We chose looking time as the dependent variable by referring to Gurler et al. (2015). In this study, Gurler et al. used looking time as the dependent variable and found that mouth-looking time was positively correlated with McGurk effect. Moreover, looking time was widely used in previous studies to reflect participants' attention to a specific AOI (e.g., Chawarska & Shic, 2009; Tsang et al., 2022).

Behavioral data anal sis. We analyzed the incongruent trials and excluded congruent trials as they were used as filler trials. For the incongruent trials, children made three types of responses: auditory responses "ba," visual responses "ga," and fused responses "da" (McGurk response). By referring to Stevenson et al. (2014), we took the fused response "da" as the McGurk percept. We computed children's percentages of each type of response in both conditions. We conducted the following analyses using non-parametric statistical analyses (i.e., repeated measures permutation ANOVA, Wilcoxon signed ranks tests, Mann-Whitney *U*-test) as the data violated the normal distribution.

2.2 | Results

2.2.1 | Blurring eyes decreased eyes-looking time and increased mouth-looking time

To explore whether blurring eyes could change looking time in the two groups on the speaker's eyes and mouth, we conducted a 2 × 2 × 2 repeated measures ANOVA on looking time with Condition (clear-eyes vs. blurred-eyes) and Region (eyes vs. mouth) as the within-subject factors, and Group (AC vs. NAC) as the between-subject factor using the R package "bruceR." We found a significant main effect of Condition, $F(1, 58) = 4.23$, $p = 0.04$, $\eta_p^2 = 0.07$, a significant main effect of Region, $F(1, 58) = 24.07$, $p < 0.001$, $\eta_p^2 = 0.29$, and a significant main effect of Group, $F(1, 58) = 13.13$, $p = 0.001$, $\eta_p^2 = 0.19$. It also showed a significant Region × Group interaction, $F(1, 58) = 8.73$, $p = 0.005$, $\eta_p^2 = 0.13$, and a significant Condition × Region interaction, $F(1, 58) = 33.69$, $p < 0.001$, $\eta_p^2 = 0.37$. Neither the Condition × Group interaction, $F(1, 58) = 0.05$, $p = 0.82$, $\eta_p^2 = 0.001$, nor the Condition × Region × Group interaction, $F(1, 58) = 0.07$, $p = 0.80$, $\eta_p^2 = 0.001$, was significant.

For the significant Condition × Region interaction, we further conducted simple effect analyses to test the condition difference of children's looking time on the eyes and the mouth. We found that children's looking time was significantly different between the clear-eyes condition and blurred-eyes condition for both the eyes, $F(1, 58) = 31.83$, $p < 0.001$, $\eta_p^2 = 0.35$, and the mouth, $F(1, 58) = 15.25$, $p < 0.001$, $\eta_p^2 = 0.21$. The significant difference in the eyes indicates that the eyes-looking time of the two groups decreased in the blurred-eyes condition ($M_{AC} = 0.25$, $SD_{AC} = 0.17$; $M_{NAC} = 0.24$, $SD_{NAC} = 0.23$; AC for autistic group and NAC for nonautistic group) compared to the clear-eyes condition ($M_{AC} = 0.48$, $SD_{AC} = 0.36$; $M_{NAC} = 0.46$, $SD_{NAC} = 0.36$; see Figure 3a). The significant difference in the mouth area indicates that the mouth-looking time of the two groups increased in the blurred-eyes condition ($M_{AC} = 0.55$, $SD_{AC} = 0.36$; $M_{NAC} = 0.85$, $SD_{NAC} = 0.32$) compared to the clear-eyes condition ($M_{AC} = 0.39$, $SD_{AC} = 0.24$; $M_{NAC} = 0.72$, $SD_{NAC} = 0.35$; see Figure 3b).

For the significant Region × Group interaction, we also conducted simple analyses to test the group difference of the two groups' looking time on the eyes and the mouth. The results showed that the autistic group and nonautistic group spent similar time viewing the eyes, $F(1, 58) = 0.06$, $p = 0.81$, $\eta_p^2 = 0.001$, but significantly different time viewing the mouth, $F(1, 58) = 17.57$, $p < 0.001$, $\eta_p^2 = 0.23$. The significant difference in mouth-looking time indicates that the autistic group ($M_{blur} = 0.55$, $SD_{blur} = 0.36$; $M_{clear} = 0.39$, $SD_{clear} = 0.24$) spent significantly less time viewing the mouth than the nonautistic group ($M_{blur} = 0.85$, $SD_{blur} = 0.32$; $M_{clear} = 0.72$, $SD_{clear} = 0.35$) did. We also explored whether blurring eyes changed children's looking time on the nose and the other area. Results only showed a significant effect of group for both areas, please see the Supplemental materials for detail (see Figure S1).

2.2.2 | Blurring eyes enhanced the McGurk effect in autism

We further tested the group difference of the three kinds of responses (i.e., auditory responses "ba," visual responses "ga," and McGurk responses "da") in the clear-eyes and blurred-eyes condition separately and found that the autistic group showed less McGurk effect than the nonautistic group in both conditions (see Figure S2 in Supplemental materials for detailed information).

To examine the condition and group differences of the McGurk effect ("da" response), we performed a two-way repeated measures permutation ANOVA with Condition (clear-eyes vs. blurred-eyes) as the within-subject factor and Group (AC vs. NAC) as the between-subject factor using the *R* package "permuco" default method (Frossard & Renaud, 2019; R Core Team, 2017). The results showed a significant main effect of Group, $F(1, 58) = 25.56$, permutation $p = 0.0002$, $\eta_p^2 = 0.31$, and a significant Group × Condition interaction, $F(1, 58) = 5.82$, permutation $p = 0.02$, $\eta_p^2 = 0.09$, but no main effect of Condition, $F(1, 58) = 1.80$, permutation $p = 0.19$, $\eta_p^2 = 0.03$. We further conducted a Wilcoxon signed-rank test to examine the differences in the McGurk effect for each group. The results showed that the autistic group had a stronger McGurk effect, $z = 2.53$, $p = 0.01$, $r = 0.46$, in the blurred-eyes condition than in the clear-eyes condition, and that the nonautistic group showed a similar McGurk effect in two conditions, $z = 0.92$, $p = 0.36$, $r = 0.17$ (see Figure 4).

In summary, for AC, blurring the speaker's eyes decreased their eyes-looking time, increased their mouth-looking time, and increased their audiovisual speech integration in the McGurk task compared with the clear-eyes condition. For NAC, blurring the speaker's eyes did not change their audiovisual speech integration in the McGurk task, although their eyes-looking time was decreased and mouth-looking time was increased compared with the clear-eyes condition.

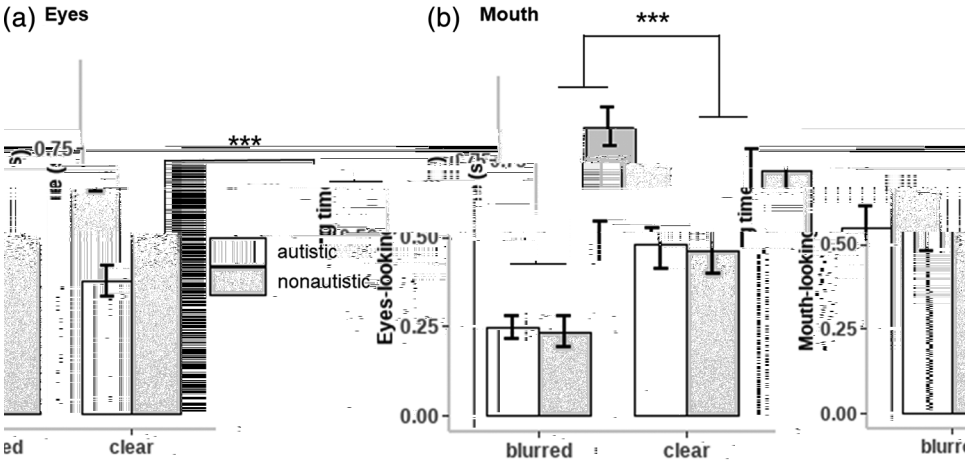

FIGURE 3 Eyes-looking time and mouth-looking time in Experiment 1. *Note.* Eyes-looking time (a) and mouth-looking time (b) in the autistic and nonautistic groups in the clear-eyes condition and blurred-eyes condition in Experiment 1. Error bars represent SEMs. ***$p < 0.001$

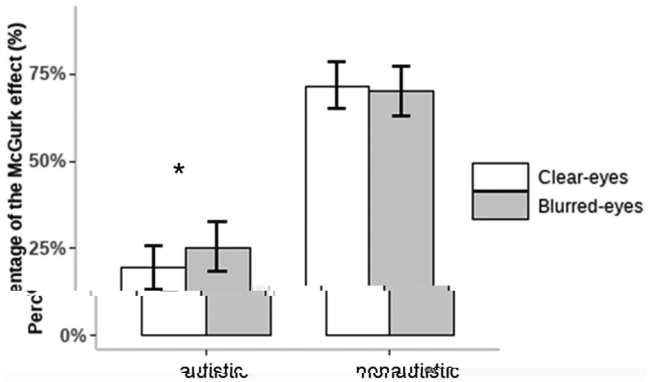

FIGURE 4 Percentage of the McGurk effect in Experiment 1. *Note.* Percentage of the McGurk effect in the autistic and nonautistic groups under the two conditions in Experiment 1. Error bars represent SEMs. *$p < 0.05$

3 | EXPERIMENT 2

In this experiment, to explore whether cuing children's attention to the speaker's mouth or eyes could affect children's audiovisual speech integration in the McGurk task, we compared a cue-to-mouth condition, a cue-to-eyes condition, and a free-viewing condition. In these three conditions, we measured children's audiovisual speech perception employing the McGurk paradigm and tracked their eye movements.

3.1 | Method

3.1.1 | Participants

Forty AC (age range: 4.28-7.18 years; all boys) and 42 NAC (age range: 4.60-7.35 years; all boys) took part in the present study. AC were from a specialized school, and NAC were from a kindergarten and an elementary school. Identical to Experiment 1, all AC were diagnosed according to the DSM-V criteria, and their diagnoses were further confirmed by the Chinese version of the AQ-Child (American Psychiatric Association, 2013; Auyeung et al., 2008). The two groups were matched for both age and IQ (see Table 1 for detailed information). IQ was also measured using the Chinese version of the WPPSI-IV (Wechsler, 2014). Among all participants, 25 AC (age range: 4.55-6.82 years; all boys) and 26 NAC (age range: 4.95-6.79 years; all boys) participated in both Experiment 1 and Experiment 2. For these participants, the two experiments were completed on the same day and Experiment 2 was completed after Experiment 1 and a short break. The experiment was approved by the research ethics committee of Peking University. Before the experiment, parents of all children provided written informed consent.

3.1.2 | Stimuli and procedures

The stimuli and apparatus used in this experiment were identical to those in the clear-eyes condition in Experiment 1. The present experiment also included a practice session and a formal session. The formal session began with eye movement calibration and included three conditions: cue-to-mouth, cue-to-eyes, and free-viewing. Each condition consisted of four trials of congruent "ba," four trials of congruent "ga," and 12 trials of incongruent "AbVg" (auditory "ba" + visual "ga"). In the cue-to-mouth condition, each trial began with a black screen with an oval at the position where the speaker's mouth would appear, and children were directed to look at the oval (see Figure 1b). If the children fixated on the oval area for at least 500 ms, the stimulus was presented. Finally, a black screen was displayed until the children responded. Children's responses were recorded by the experimenter's pressing of "b," "d," and "g" on the keyboard for responses of "ba," "da," and "ga" respectively. In the cue-to-eyes condition, the procedure was identical to that in the cue-to-mouth condition, except that the oval was present at the position where the speaker's eyes would appear. In the free-viewing condition, no oval was presented, and the stimulus was not displayed

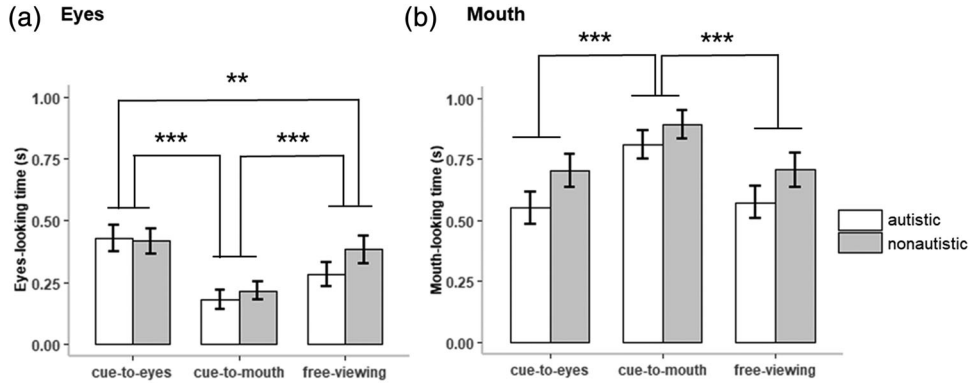

FIGURE 5 Eyes-looking time and mouth-looking time in Experiment 2. *Note*. Eyes-looking time (a) and mouth-looking time (b) in the autistic and nonautistic groups under the three conditions in Experiment 2. Error bars represent SEMs. $^{**}p < 0.01$. $^{***}p < 0.001$

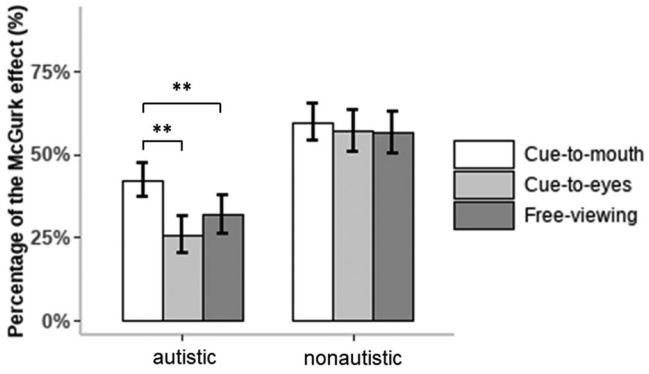

FIGURE 6 Percentage of the McGurk effect in Experiment 2. *Note*. Percentage of the McGurk effect in the three conditions for the autistic and nonautistic groups in Experiment 2. Error bars represent SEMs. $^{**}p < 0.01$

In sum, for AC, cuing to the mouth increased their mouth-looking time and increased their audiovisual speech integration in the McGurk task compared with the other two conditions—"Cue-to-eyes" and "Free-viewing." For NAC, cuing to the mouth did not change their audiovisual speech integration in the McGurk task, although their mouth-looking time also increased compared with the other two conditions.

4 | GENERAL DISCUSSION

In the present study, we aimed to increase audiovisual speech integration in the McGurk task in AC by increasing their mouth-looking time. In two experiments, we managed to increase the mouth-looking time by blurring the speaker's eyes (Experiment 1) and cuing children's first fixation to the speaker's eyes or mouth (Experiment 2). We found that blurring the speaker's eyes and cuing the first fixation to the speaker's mouth could enhance audiovisual speech integration in the McGurk task in AC. At the same time, we found that blurring the speaker's eyes and cuing the first fixation to the speaker's mouth also increased the mouth-looking time in NAC, but did not enhance the audiovisual speech integration in the McGurk task in NAC.

First, blurring the speaker's eyes and cuing children's first fixation to the speaker's mouth increased the mouth-looking time and increased audiovisual speech integration in the McGurk task in AC. These findings confirmed our hypotheses. Previous studies found that mouth-looking time positively correlated with audiovisual speech integration in the McGurk task in AC (Feng et al., 2021) and nonautistic adults (Gurler et al., 2015). To increase audiovisual speech integration in the McGurk task in AC, we adopted two manipulations in two experiments to increase their mouth-looking time—blurring the speaker's eyes and cueing children's first fixation to the speaker's mouth. We found that these two manipulations increased the mouth-looking time in AC, as expected. At the same time, we also found that these two manipulations enhanced audiovisual speech integration in the McGurk task in AC. Our findings extend the previous evidence on the relationship between mouth-looking time and audiovisual speech integration in the McGurk task in AC (Feng et al., 2021) by further revealing that audiovisual speech integration in the McGurk task in AC could be increased by increasing their mouth-looking time. Our findings not only deepen the understanding of the underlying mechanisms of audiovisual speech integration in the McGurk task in autism, but also provide important insights for supporting strategies targeting audiovisual speech integration in AC. In addition, our findings confirmed the role of visual information for speech perception in AC (Newman et al., 2021).

Second, for NAC, both blurring the speaker's eyes and cuing to the speaker's mouth increased the mouth-looking time but did not enhance their audiovisual speech integration in the McGurk task. This finding was consistent with a previous study in NAC, which found that audiovisual speech integration in the McGurk task in NAC did not correlate with mouth-looking time but correlated with eyes-looking time (Feng et al., 2021). However, another study in adults found that audiovisual speech integration correlated with mouth-looking time (Gurler et al., 2015). These controversies indicate that the relationship between audiovisual speech integration in the McGurk task and visual attention to different core facial features could vary with age. One possible

the underlying mechanisms of audiovisual speech integration in autism. This finding could also provide insights for the development of supports to increase audiovisual speech integration in AC.

ACKNOWLEDGMENTS

The authors are grateful to Zipeng Ma, Fuli Liu, Yanhong Wu, Yinan Lv, and the staff in Qingdao Elim School, for their generous assistance in completing the study. This work was supported by Key-Area Research and Development Program of Guangdong Province (2019B030335001), National Natural Science Foundation of China (31871116, 62061136001, 31861133012, 32271116), Philosophy and Social Science Foundation of Hunan Province (19YBQ109), Outstanding Youth Foundation of Education Department of Hunan Province (21B0004), and Natural Science Foundation of Hunan Province of China (2022JJ40580).

CONFLICT OF INTEREST

The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

Data were available upon reasonable request.

ETHICS APPROVAL STATEMENT

The experiment was performed in compliance with the institutional guidelines set by the Ethics Committee of School of Psychological and Cognitive Sciences, Peking University, China, in accordance to the 1975 Declaration of Helsinki concerning human and animal rights.

REFERENCES

Alsius, A., Navarra, J., Campbell, R., & Soto-Faraco, S. (2005). Audiovisual integration of speech falters under high attention demands. *Current Biology*, *15*(9), 839–843. https://doi.org/10.1016/j.cub.2005.03.046

Alsius, A., Navarra, J., & Soto-Faraco, S. (2007). Attention to touch weakens audiovisual speech integration. *Experimental Brain Research*, *183*, 399–404. https://doi.org/10.1007/s00221-007-1110-1

Altieri, N., Pisoni, D. B., & Townsend, J. T. (2011). Some behavioral and neurobiological constraints on theories of audiovisual speech integration: A review and suggestions for new directions. *Seeing and Perceiving*, *24*(6), 513–539. https://doi.org/10.3389/fpsyg.2014.00257

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

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

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*, *14*, 2592–2602. https://doi.org/10.1002/aur.2598

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

Grynszpan, O., Simonin, J., Martin, J. C., & Nadel, J. (2012). Investigating social gaze as an action-perception online performance. *Frontiers in Human Neuroscience*, *6*(6), 94. https://doi.org/10.3389/fnhum.2012.00094

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

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

Jones, J., & Callan, D. E. (2003). Brain activity during audiovisual speech perception: An fMRI study of the McGurk effect. *NeuroReport*, *14*(8), 1129–1133.

Kleiner, M., Brainard, D., & Pelli, D. (2007). What's new in Psychtoolbox-3? *Perception*, *36*, ECVP Abstract Supplement.

Lewkowicz, D. J., & Hansen-Tift, A. M. (2012). Infants deploy selective attention to the mouth of a talking face when learning speech. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(5), 1431–1436.

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

Mercier, M. R., & Cappe, C. (2022). The interplay between multisensory integration and perceptual decision making. *NeuroImage*, *222*, 116970.

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 Disorder*, *38*, 1349–1358. https://doi.org/10.1007/s10803-007-0521-y

Moriuchi, J. M., Klin, A., & Jones, W. (2017). Mechanisms of diminished attention to eyes in autism. *American Journal of Psychiatry*, *174*(1), 26–35. https://doi.org/10.1176/appi.ajp.2016.15091222

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