

RESEARCH ARTICLE

Staged Audiovisual Speech Integration and Altered Early-Stage Audiovisual Processing in Autistic Children: An EEG Investigation

Shuyuan Feng^{1,2} | Baolin Li³  | Wei Ni² | Sio Pan Hoi² | Xue Li⁴ | Juan Zhang^{5,6} | Qiangdong Wang⁷  | Lihan Chen^{2,8,9}  | Li Yi^{2,9,10} 

¹Institute for Applied Linguistics, School of Foreign Languages, Central South University, Changsha, Hunan, China | ²School of Psychological and Cognitive Sciences, Peking University, Beijing, China | ³School of Psychology, Shaanxi Normal University, Xi'an, Shaanxi, China | ⁴Peking University Sixth Hospital, Beijing, China | ⁵Faculty of Education, University of Macau, Macau, China | ⁶Centre for Cognitive and Brain Sciences, University of Macau, Macau, China | ⁷Faculty of Psychology, Beijing Normal University, Beijing, China | ⁸National Engineering Laboratory for Big Data Analysis and Applications, Peking University, Beijing, China | ⁹Key Laboratory of Machine Perception (Ministry of Education), Beijing, China | ¹⁰IDG/McGovern Institute for Brain Research at PKU, Peking University, Beijing, China

Correspondence: Li Yi (yilipku@pku.edu.cn) | Lihan Chen (clh@pku.edu.cn)

Received: 13 September 2025 | **Revised:** 3 April 2026 | **Accepted:** 4 May 2026

Keywords: audiovisual incongruence | audiovisual speech integration | autism | EEG | McGurk effect

ABSTRACT

Autistic children exhibit difficulties in audiovisual speech integration, which are associated with their social communication challenges. The neural mechanism underlying audiovisual speech integration difficulties in autism remains unclear. We recruited 19 neurotypical (NT) children and 29 autistic children. We recorded their behavioral responses and Electroencephalography (EEG) signals to audiovisual congruent syllables, and incongruent syllables that could evoke audiovisual speech integration (i.e., McGurk effect). For the EEG analysis, we further classified autistic children into the autistic McGurk group and the autistic non-McGurk group based on their strength of audiovisual speech integration. Behaviorally, we found reduced audiovisual speech integration in autistic children. At the neural level, we found that: (1) NT children showed an early stage audiovisual processing (indexed by the N1 amplitude), which was altered in two autistic groups; (2) all three groups exhibited a successful audiovisual incongruence detection (i.e., phonological Mismatch Negativity, pMMN); (3) NT and autistic McGurk groups could successfully resolve the audiovisual incongruence (indexed by the restoration of pMMN), but the autistic non-McGurk group could not (indexed by the sustained negative amplitude). Furthermore, we found a distinct temporal decoding pattern between non-McGurk and congruent trials across groups: both NT and autistic McGurk groups exhibited early EEG decoding, whereas the autistic non-McGurk group demonstrated successful decoding during late processing stages. Audiovisual speech integration entails a three-stage process in NT children: early audiovisual processing, audiovisual incongruence detection, and audiovisual incongruence resolution. The altered early-stage processing was possibly the neural mechanism underlying the reduced audiovisual speech integration in autistic children.

Shuyuan Feng and Baolin Li contributed equally to this study.

© 2026 International Society for Autism Research and Wiley Periodicals LLC.

Summary

By employing EEG technique, we explored the neural mechanism underlying audiovisual speech integration difficulties in autism. We found that autistic groups showed an altered early-stage audiovisual processing (indexed by an altered N1 amplitude) compared to NT children. NT group demonstrated a three-stage process in audiovisual speech integration: early response to audiovisual incongruency, audiovisual incongruence detection, and audiovisual incongruence resolution. The altered early-stage audiovisual processing might be seen as a neural signature of autism in audiovisual speech processing, potentially contributing to the diagnosis of autism.

1 | Introduction

Autism is a neurodevelopmental condition characterized by altered social interaction and communication, as well as restricted and repetitive behaviors (DSM-5; American Psychiatric Association 2013). Autistic children showed reduced audiovisual speech integration compared to neurotypical (NT) children as reported in previous studies (e.g., Bebko et al. 2014; Feng et al. 2021, 2023; Stevenson et al. 2014). Audiovisual speech integration refers to the process that combines a speaker's auditory speech signals and facial articulatory movements, forming unified perceptual representations (Altieri et al. 2011). Reduced audiovisual speech integration is associated with decreased language and communication abilities in autistic individuals (Feldman et al. 2018). Unveiling the neural mechanism underlying reduced audiovisual speech integration in autistic children provides insights for the mechanisms of their delayed language development and social communication challenges.

In NT individuals, the neural mechanisms underlying audiovisual speech integration have been explored using McGurk tasks (Proverbio et al. 2018; Roa Romero et al. 2015). The McGurk effect occurs when an auditory phoneme is dubbed onto the articulatory movement of another phoneme, resulting in the fused perception of a third phoneme (McGurk and MacDonald 1976). In NT infants and children, to the best of our knowledge, only Kushnerenko et al. (2008) explored NT infants' neural response to McGurk stimuli. They found that 5-month-old infants demonstrated similar event-related brain potentials (ERPs) when processing audiovisual congruent stimuli and McGurk stimuli. This pattern was different from those when processing audiovisual incongruent stimuli that cannot be integrated. However, this study did not identify any specific components for McGurk stimuli processing in NT infants.

In NT adults, several studies investigated the neural mechanisms underlying audiovisual speech integration (e.g., Proverbio et al. 2018; Roa Romero et al. 2015). Using EEG technique, Proverbio et al. (2018) identified that audiovisual incongruent syllables (McGurk stimuli) evoked a phonological Mismatch Negativity (pMMN) in NT adults about 500 ms after the voice onset compared to audiovisual congruent syllables. They defined this component as electrophysiological evidence of automatic detection of incongruence between auditory phonemes and visual articulatory movements in McGurk stimuli. In

another study, Roa Romero et al. (2015) found a reduced N1 component (78–170 ms), an early (50–250 ms) and a late (500–620 ms) suppression of beta-band power in McGurk trials compared to audiovisual congruent trials in NT adults. On the basis of these findings, they proposed a possible three-stage process in audiovisual speech integration, including (a) early enhanced integrative processing (i.e., N1 reduction: 78–170 ms), (b) audiovisual incongruence detection and allocation of processing demands (i.e., early beta power suppression: 50–250 ms), and (c) audiovisual incongruence resolution and formation of a coherent perception (i.e., late beta power suppression: 500–620 ms; Roa Romero et al. 2015).

In autistic children, a series of studies revealed that autistic children showed reduced audiovisual speech integration compared to NT children, manifesting by their lower ability of McGurk perception (e.g., Bebko et al. 2014; Feng et al. 2021, 2023). This reduced audiovisual speech integration in autistic children suggests that they process McGurk stimuli differently from NT children at the neural level. However, the underlying mechanism has not been fully explored. Thus, the present study aimed to unveil the neural mechanism underlying audiovisual speech integration difficulties in autistic children through EEG techniques. Building upon the three-stage proposition and subsequent empirical evidence (e.g., N1, pMMN; Proverbio et al. 2018; Roa Romero et al. 2015), we proposed a modified three-stage process in audiovisual speech integration, specifically addressing the dynamic interplay among the three stages. The first stage would be an early response to audiovisual incongruency indexed by the N1 amplitude. The second stage would be the audiovisual incongruency detection indexed by the pMMN. The third stage would be the audiovisual incongruence resolution indexed by restoration of pMMN's negative deflection.

To test our proposition, we employed the McGurk effect paradigm to explore the neural activities in autistic and NT children in the present study. We divided the autistic children into autistic McGurk group and autistic non-McGurk group based on their McGurk performance. Then, we compared the neural activities between the processing of McGurk/non-McGurk trials and audiovisual congruent trials in each group. We also analyzed whether EEG signals could decode different kinds of trials in each group employing Multivariate Pattern Analysis (MVPA). At the behavioral level, we hypothesized that autistic children showed reduced McGurk perception than NT children. For the neural activities, we hypothesized that (a) the NT group would demonstrate a modified three-stage process in the neural processing of McGurk/non-McGurk trials, but the two autistic groups might demonstrate at least one alteration in the three-stage process; (b) the three groups might show some differences in neural decoding performance.

2 | Methods

2.1 | Participants

Nineteen NT boys (age: 6.22–14.90 years, $M = 9.84$ years, $SD = 2.10$) and 29 autistic boys (age: 6.14–15.59 years, $M = 9.16$ years, $SD = 2.41$) participated in the present study. All participants were Mandarin-speaking Chinese and were recruited from urban areas of Beijing,

China. All autistic boys had received a diagnosis of autism based on the criteria of the Diagnostic and Statistical Manual of Mental Disorders (DSM-V; American Psychiatric Association 2013). Autism diagnosis was further confirmed by the Chinese version of the Autism Spectrum Quotient: Children's Version (AQ-Child; Auyeung et al. 2008). The two groups were matched in both age and IQ, measured by the Chinese version of the Wechsler Intelligence Scale for Children-Fourth Edition (WISC-IV CN; Wechsler 2014; see Table 1 for detailed information). This study was approved by the research ethics committee of Peking University. We obtained oral consent from all children and written consent from their parents before the experiment.

All participants' data were included in the behavioral data analysis. In the EEG data analysis, we conducted a cluster analysis to classify autistic children into two subgroups based on their percentages of the McGurk effect using k-means clustering ($k=2$). This analysis identified two distinct clusters: the high McGurk effect cluster (McGurk effect $\geq 45\%$; $N=15$ in autistic McGurk group) and the low McGurk effect cluster (McGurk effect $\leq 10\%$; $N=14$ in autistic non-McGurk group). The two clusters corresponded to the autistic McGurk group and autistic non-McGurk group in the present study. In the NT group, most children showed a McGurk effect of higher than 45% and only three children showed very low McGurk effect (about 10%). Due to the small sample size, these three children could not form a group and were therefore excluded from the EEG data analysis. Thus, in the NT group, all the remaining 16 children who showed more than 45% of McGurk effect were taken into the EEG data analysis. We conducted a power analysis using G*Power (Faul et al. 2009). The analysis indicated that a total sample size of $N=42$ was required for a design involving three groups (each with two conditions: congruent vs. incongruent) assuming $\alpha=0.05$, power=0.80, and a medium effect size (Cohen's $f=0.25$). Thus, the current sample size ($N=45$) was considered adequate. In addition, the sample sizes in each group are comparable to those reported in the previous EEG study by Roa Romero et al. (2016). The autistic McGurk group, the autistic non-McGurk group, and the NT group were matched in both age and IQ (see Table 1, "EEG analysis" for detailed information).

2.2 | Stimuli

We adopted the McGurk effect to measure participants' audiovisual speech integration (McGurk and MacDonald 1976). We

employed two types of stimuli: audiovisual congruent stimuli and audiovisual incongruent stimuli. In these two types of stimuli, a female speaker was videotaped while articulating syllables. The audiovisual congruent stimuli were the original videos of the speaker articulating "ba" and "da". The audiovisual incongruent stimuli were obtained by dubbing the visual "ga" onto the auditory "ba" (auditory "ba" + visual "ga") from the original videos using Adobe Premiere Software Pro CS 6.0. The audiovisual incongruent stimuli could evoke the McGurk illusory percept of "da", which indicates the occurrence of the audiovisual speech integration.

All the stimuli were videos with a resolution of 1280×720 pixels and a frame rate of 25 frames/s. Each stimulus lasted approximately 1.85 s. All videos began and ended with the speaker's still face. For the audiovisual congruent stimuli "ba" and "da", the sounds began after the speaker opened her mouth 160 and 200 ms respectively. For the audiovisual incongruent stimuli, the sounds began after the speaker opened her mouth 240 ms. We obtained written consent from the female speaker to use these videos in the experiment and publications.

2.3 | Procedures

The experiment was conducted in an electrically shielded room with dimmed lights. Participants were seated before the 21.5-inch display screen with a resolution of 1920×1080 pixels. In the center of the screen, the stimuli were presented with Psychtoolbox (Brainard 1997; Kleiner et al. 2007; Pelli 1997) in MATLAB 2016b (The MathWorks, Natick, MA). Sounds were presented through two speakers located on the two sides of the screen.

We implemented a practice session and the formal experiment. The practice session was implemented to familiarize participants with the McGurk task. The formal experiment included 10 blocks. Each block included two conditions: the audiovisual congruent condition and the audiovisual incongruent condition. In the audiovisual congruent condition, six audiovisual congruent "ba" and six audiovisual congruent "ga" were presented. In the audiovisual incongruent condition, 12 audiovisual incongruent "AbVg" were presented. Totally, each block included 24 trials, which were presented in random order. For a typical trial, the procedure was shown in Figure 1. First, a fixation was presented on the screen for

TABLE 1 | Participants' characteristics in the behavior analysis and the EEG analysis.

	Group	N	Male/female	Mean age in years (SD)	IQ (WISC-IV)
Behavior analysis	Autism	29	29/0	9.16 (2.41)	109.03 (18.61)
	NT	19	19/0	9.84 (2.10)	118.42 (13.70)
	<i>t</i> (<i>p</i>)	N/A	N/A	-1.04 (0.31)	-1.89 (0.07)
EEG analysis	Autistic McGurk	15	15/0	9.49 (3.10)	105.13 (22.32)
	Autistic non-McGurk	14	14/0	8.80 (1.40)	113.21 (13.15)
	NT_Mc	16	16/0	10.03 (2.12)	119.81 (13.12)
	<i>F</i> (<i>p</i>)	N/A	N/A	1.04 (0.36)	2.97 (0.06)

1000 ms. Then, the first static image of a video was presented briefly with a random duration of 400 ms, 600 ms, or 800 ms to minimize participants' expectancy effects and to control the influence of visual ERPs due to picture onset (Roa Romero et al. 2015). After that, the video was displayed. Finally, a response screen with two choices (“ba” and “da”) was presented to prompt the participants to issue the response (“1” or “3” for “ba” or “da”). The mappings between the numbers on the keyboard and the choices on the screen were counterbalanced among participants. Participants took breaks for 2–3 min after completing a block. The experiment lasted about 50 min.

2.4 | EEG Recording and Preprocessing

2.4.1 | EEG Recording

In the formal experiment, continuous EEG was recorded by EGI (Electrical Geodesic Inc., Eugene, OR) with a 128-channel HydroCel Geodesic Sensor Net, which was referenced online to the vertex electrode (Cz). Eye blinks and vertical eye movements were examined with electrodes located below and above the eyes (E8, E126, E25, E127). Horizontal eye movements were measured using E125 and E128, located at positions to the left and right of the eyes. Data were recorded at 1000 Hz with an online 0.01–50 Hz band-pass filter. Electrode impedance was kept below 50 k Ω during data collection.

2.5 | Data Analysis

2.5.1 | Behavioral Data Analysis

We scored children's responses in the audiovisual congruent condition and the audiovisual incongruent condition separately. In the audiovisual congruent condition, for the stimuli “ba”, we took the response “ba” as correct response and the response “da” as incorrect response; for the stimuli “da”, we took the response “da” as correct response and the response “ba” as incorrect response. In the audiovisual incongruent condition, we took the

response “da” as McGurk percept and the response “ba” as non-McGurk percept. We analyzed the data using nonparametric statistical analysis (i.e., Mann–Whitney U test and permutation ANOVA) as the normal distribution assumption required by parametric tests was violated.

2.5.2 | EEG Data Analysis

2.5.2.1 | EEG Pre-Processing. Offline EEG data were preprocessed using the Maryland Analysis of Developmental EEG (MADE) Pipeline (Debnath et al. 2020), a standardized EEG pipeline based on the EEGLab Toolbox for Matlab (Delorme and Makeig 2004). Following the standard procedure of preprocessing with the MADE, continuous data were down-sampled to 250 Hz, and the data from 24 outermost rings of electrodes (E17, E38, E43, E44, E48, E49, E113, E114, E119, E120, E121, E125, E126, E127, E128, E56, E63, E68, E73, E81, E88, E94, E99, and E107) were removed. Then, the data were high-pass filtered at 0.3 Hz and low-pass filtered at 40 Hz. The FASTER plugin for EEGLab (Nolan et al. 2010) was used to identify bad channels (Number of Bad Channels: Mean = 5.11, SD = 3.07, Range = 0–13). To identify artifacts in the data, we performed independent components analysis (ICA) on a copy of the dataset. The copied dataset was filtered with a 1 Hz high-pass filter and segmented into arbitrary 1000 ms epochs. Prior to ICA decomposition, epochs were removed if the voltage amplitude exceeded $\pm 1000 \mu\text{V}$ or if power in the 20–40 Hz band (after Fourier analysis) was greater than 30 dB to remove the EMG-like activity. If more than 20% of epochs in a channel were rejected, this channel was removed from both the ICA-copied dataset and the original dataset. Next, ICA was performed on the copied dataset and the ICA weights were copied back to the original continuous dataset. The “adjusted-ADJUST” Matlab scripts (Leach et al. 2020) were used to automatically identify artifactual independent components in the original dataset, which were then removed from the data. Then, epochs from –100 to 800 ms relative to the onset of the auditory stimuli were extracted for the EEG analysis. All epochs were baseline-corrected with a 100 ms

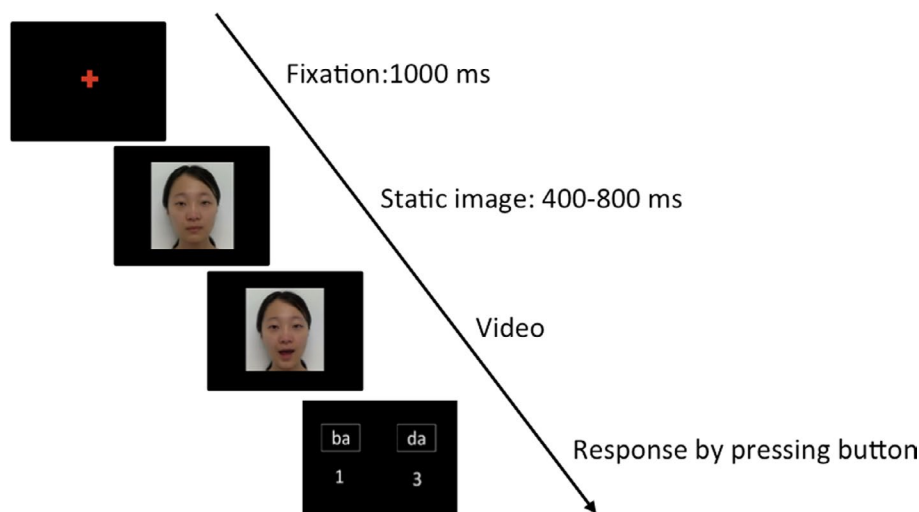


FIGURE 1 | The procedure of a sample trial. Please refer to the text for detailed information.

window (−100 to 0 ms before the onset of the auditory stimuli). A rejection threshold of $\pm 125 \mu\text{V}$ based on frontal electrodes (E1, E8, E14, E21, E25, E32) was used to further reject epochs with any ocular artifacts. For the remaining epochs, channels with voltage that exceeded the $\pm 125 \mu\text{V}$ threshold were interpolated at the channel level. However, when more than 10% of the channels exceeded this threshold, the epoch was removed. Finally, all rejected channels were interpolated with the spherical spline interpolation procedure, and data were re-referenced to the average of all electrodes. The online reference Cz was added back into the dataset.

To ensure there were enough trials in the EEG data analysis, we focused on the comparison of ERPs between congruent trials with correct responses and McGurk trials (i.e., trials evoked McGurk effect) in NT and autistic McGurk groups, and between congruent trials with correct responses and non-McGurk trials (i.e., trials did not evoke McGurk effect) in the autistic non-McGurk group. The number of trials was equalized according to the lower trials number of both types of trials in each group. There were 95 trials ($SD = 18.2$) in the NT group, 88 trials ($SD = 26.6$) in the autistic McGurk group, and 97 trials ($SD = 26.9$) in the autistic non-McGurk group on average for each condition in the EEG data analysis. There was no significant difference on the number of trials among the three groups, $F(2, 44) = 0.497, p = 0.612$.

2.5.2.2 | ERP Analysis. The ERP analyses mainly focused on the auditory ERP components in the central area (Knowland et al. 2014). To increase the signal-to-noise ratio, we averaged six electrodes around and including the apex, Cz (Cz, E7, E106, E31, E80, E55), which showed the clearest auditory components in the present study. Building on previous studies reporting neural differences between McGurk trials and AV congruent trials, we examined the ERP differences between McGurk trials and audiovisual congruent trials in NT and autistic McGurk groups (e.g., Proverbio et al. 2018; Roa Romero et al. 2015; Saint-Amour et al. 2007; Stekelenburg and Vroomen 2012). This comparison aimed to explore the neural mechanisms of audiovisual speech integration in both NT and autistic children. We further examined the ERP differences between non-McGurk trials and audiovisual congruent trials in the autistic non-McGurk group. This comparison aimed to explore, when integration failed for incongruent stimuli, whether their neural responses were similar to those elicited by audiovisual congruent stimuli. To explore the group differences, we also examined the ERP difference wave in those trials between the two groups that showed the McGurk effect (NT group and the autistic McGurk group) and between the two autistic groups (autistic McGurk group and autistic non-McGurk group). We employed the cluster-based permutation test in the comparisons with 1000 times repetition of the permutation procedure (Maris and Oostenveld 2007).

2.6 | Multivariate Pattern Analysis (MVPA)

In order to compare and contrast the patterns of neural activity in the two conditions in each group, we tested whether it is possible to decode the audiovisual congruent trials with correct

responses and audiovisual incongruent trials through MVPA of the EEG signals. The Amsterdam Decoding and Modeling toolbox (ADAM, version 1.13-beta; Fahrenfort et al. 2018) in Matlab was used to perform the MVPA. To avoid overfitting, only 21 channels from the standard EEG montage 10–20 system were included in the decoding analysis. The EEG amplitudes at individual electrode channels were used as decoding features. This resulted in 21 features in the predicted classes. Prior to the MVPA, the EEG data was down-sampled to 125 Hz to facilitate the decoding process. The main analysis involved a backward decoding classification with a linear discriminant classifier. The classifier was trained and tested on each time point with an 8-fold cross-validation: trials were first randomly divided into 8 equal-sized folds, and then the classifier was trained on 7 folds (87.5% data) and tested on the remaining fold (12.5% data). This procedure was repeated 8 times until each fold was tested. The decoding performance was averaged over folds. Two categories were involved in the analysis: congruent trials with correct responses and McGurk trials for NT and autistic McGurk groups, and congruent trials with correct responses and non-McGurk trials for autistic non-McGurk group. The Area under the Receiver Operating Characteristic (ROC) Curve (AUC) was used to assess the decoding accuracy. An AUC value of 0.5 means chance level decoding performance, and larger AUC means more accurate decoding performance. MVPA results were tested at the group level for AUC deviations from chance (i.e., 0.5). Cluster-based permutation was used as correction for multiple comparisons in the MVPA (by default 1000 iterations).

3 | Results

3.1 | Reduced McGurk Effect in Autistic Group Compared to NT Group

We first compared the McGurk effect difference between the autistic group and the NT group in the audiovisual incongruent condition using a Mann–Whitney U test by taking all children into consideration. The results showed that the autistic group showed reduced McGurk effect compared with the NT group, $U = 439.50, Z = -3.47, p = 0.001, r = -0.50$ (Figure 2a). Then, we compared the McGurk effect difference among the autistic McGurk group, the autistic non-McGurk group, and the NT group in the EEG analysis by conducting a permutation ANOVA with group as the between-subject factor using the R package “lmPerm” and “multcomp” default method (Hothorn et al. 2008; Wheeler and Torchiano 2016). Results showed a significant main effect of group, $F(2, 43) = 148.99$, permutation $p < 0.001, \eta_p^2 = 0.88$. Further, we conducted multiple comparisons using Mann–Whitney U tests to examine the differences between each pair among the three groups. We found that the autistic McGurk group showed reduced McGurk effect than the NT group did, $U = 64.50, Z = -2.22, p = 0.03, r = -0.40$, the autistic non-McGurk group also showed reduced McGurk effect than the NT group did, $U = 0.00, Z = -4.71, p < 0.001, r = -0.86$, and the autistic non-McGurk effect group showed reduced McGurk effect than the autistic McGurk group did, $U = 0.00, Z = -4.62, p < 0.001, r = -0.86$ (Figure 2b). We also analyzed children’s accuracies in the audiovisual congruent condition, which were all close to 100% in the three groups.

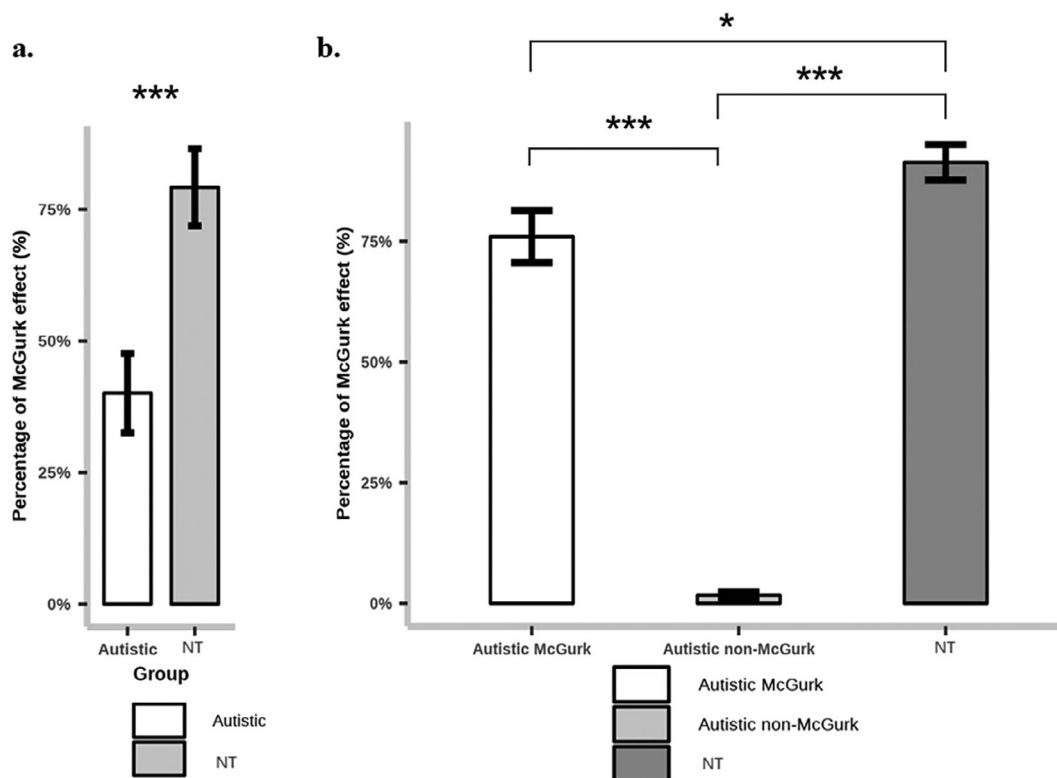


FIGURE 2 | McGurk effect in the autistic group and the NT group in the behavioral data analysis (a); and McGurk effect in the autistic McGurk group, the autistic non-McGurk group and the NT group in the EEG data analysis (b).

3.2 | Larger N1 Amplitudes Difference Between McGurk and Congruent Trials in NT Group

We compared the average central ERP between McGurk trials/non-McGurk trials and audiovisual congruent trials in the three groups by conducting a cluster-based permutation test. In the NT group, we found two significant clusters, 92–172 ms, $p=0.002$, and 420–484 ms, $p=0.012$ (Figure 3a). These two clusters respectively indicated a larger N1 amplitude and a pMMN in the McGurk trials than that in the audiovisual congruent trials. In the autistic McGurk group, we found a marginally significant cluster, 500–564 ms, $p=0.068$ (Figure 3b), which would probably be a pMMN in the McGurk trials. In the autistic non-McGurk group, we found two significant clusters, 384–672 ms, $p<0.001$, and 704–796 ms, $p=0.034$ (Figure 3c). The two clusters indicated a pMMN and a late negative component (after 700 ms) in the non-McGurk trials than that in the audiovisual congruent trials. To sum up, pMMN appeared in all three groups, and the N1 amplitudes were different across different kinds of trials in the NT group and showed no difference in the two autistic groups.

We further tested whether the three groups showed different ERP amplitude changes across different kinds of trials. We first calculated the difference waves separately for each group by subtracting the neural responses in the audiovisual congruent trials from those in the McGurk/non-McGurk trials. Then, we compared the difference waves across groups by conducting a cluster-based permutation test. Results showed

significant clusters between the NT and autistic McGurk groups, 128–148 ms, $p<0.001$ (Figure 3d) and between the autistic non-McGurk and autistic McGurk groups, 8–28 ms, $p<0.001$, and 708–748 ms, $p<0.001$ (Figure 3e). It indicated that the autistic McGurk group showed a smaller magnitude of N1 difference wave compared to the NT group and smaller magnitudes of difference wave in early processing but larger magnitudes of difference wave in late processing compared to the autistic non-McGurk group.

3.3 | Magnitudes of the N1 Difference Wave Were Correlated With the Strength of McGurk Effect in NT and Autistic McGurk Groups

As both magnitudes of N1 difference wave and strength of McGurk effect differed between the NT and the autistic McGurk groups, we analyzed the correlation between mean magnitudes of N1 difference wave and the strength of the McGurk effect in the NT and the autistic McGurk groups. Results showed that mean magnitudes of N1 difference wave (time range: 128–148 ms) were negatively correlated with the McGurk effect at the behavioral level, Spearman's $\rho=-0.508$, $p=0.004$ (Figure 4) in the NT and autistic McGurk groups. It suggests that the changes in the N1 amplitude can predict the McGurk effect in behavior in the NT and autistic McGurk groups. That is, the larger the N1 amplitude that differed across McGurk trials and audiovisual congruent trials, the greater the McGurk effect became in the two groups.

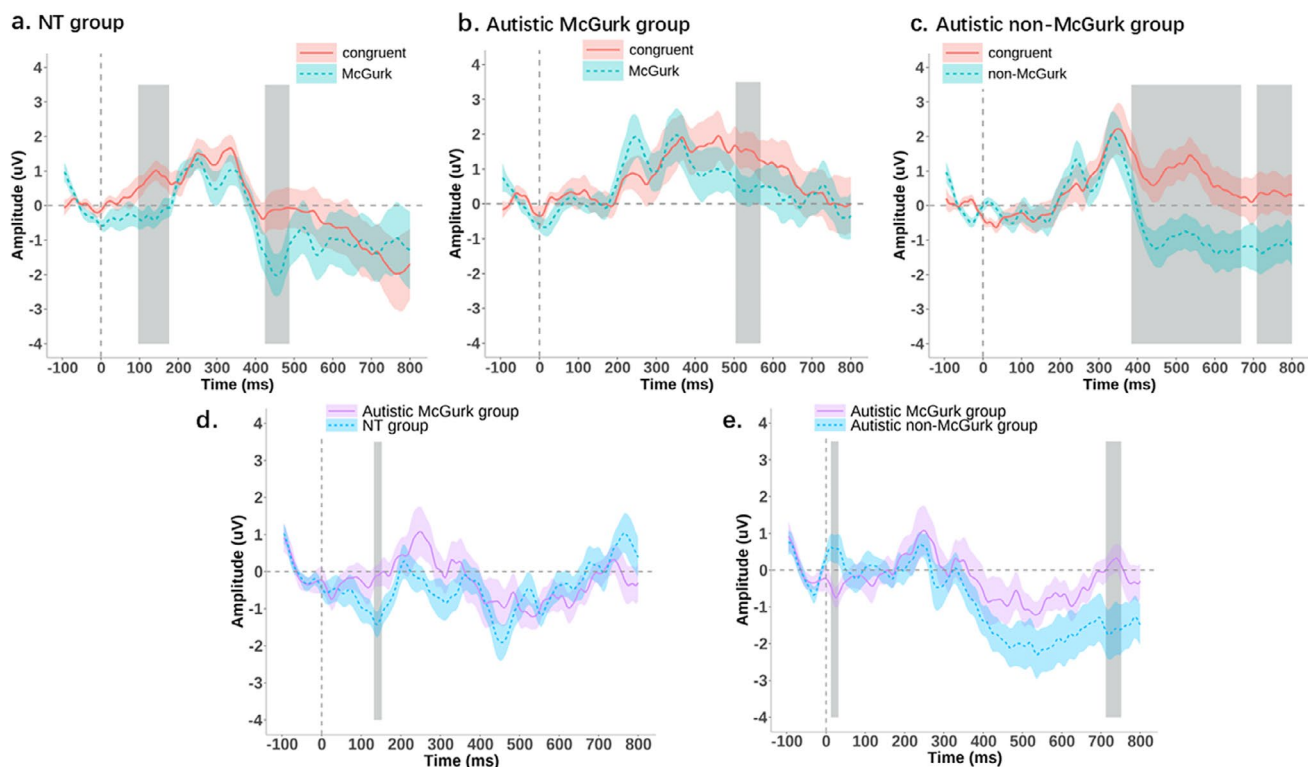


FIGURE 3 | The ERP results. Grand averaged ERPs displayed between -100 and 800 ms around the onset of auditory stimulus and recorded at the central scalp recording sites in the McGurk/non-McGurk trials and audiovisual congruent trials for the NT, autistic McGurk and autistic non-McGurk groups (a–c). The difference waves between NT and autistic McGurk groups (d) and between autistic McGurk and autistic non-McGurk groups (e). The difference waves were calculated separately for each group by subtracting the neural responses in the audiovisual congruent trials from those in the McGurk/non-McGurk trials. Significant time windows (a, c, d, and e) and marginal significant time window (b) were indicated by gray areas ($p \leq 0.05$, after cluster-based permutation).

3.4 | Superior Neural Decoding Performance in NT Group

We further explored whether the transient neural signals could decode McGurk/non-McGurk trials and audiovisual congruent trials in the three groups by conducting a MVPA. The results of the diagonal decoding are shown in Figure 5. In the NT group, above-chance decoding performances (i.e., AUC) were appeared in the entire process, which ranged from 0 to 800 ms relative to auditory stimulus onset, $p < 0.001$ (Figure 5a). In the autistic McGurk group, the above-chance decoding performances occurred roughly in the earlier stage of the process, from 16 to 420 ms, $p < 0.001$ (Figure 5b). In the autistic non-McGurk group, the above-chance decoding performances occurred roughly in the late stage of the process, which ranged from 336 to 416 ms, $p = 0.007$, and from 436 to 800 ms, $p = 0.004$ (Figure 5c).

In addition, we compared the decoding differences among the two McGurk groups and the two autistic groups. Compared to the autistic McGurk group, the NT group showed a higher decoding performance from 576 to 732 ms, $p < 0.001$ (Figure 5d), and the autistic non-McGurk group showed a higher decoding performance from 748 to 800 ms, $p = 0.008$ (Figure 5e). That is, compared to the autistic McGurk group, both the NT group and the autistic non-McGurk group classified the two kinds of stimuli better in the later stage of processing.

4 | Discussion

In the present study, we investigated the neural mechanisms underlying the audiovisual speech integration in autistic children, including the autistic McGurk group and the autistic non-McGurk group, when compared with NT children employing EEG. Behaviorally, we found that autistic children showed reduced audiovisual speech integration relative to the NT group. As for the neural activities, only the NT group demonstrated a modified three-stage process as proposed by us in the above: early response to audiovisual incongruency (i.e., an enhanced N1 amplitude), audiovisual incongruence detection (i.e., occurrence of pMMN), and audiovisual incongruence resolution (i.e., restoration of pMMN). Specifically, the three groups demonstrated the following neural activities in the three stages. In the first stage, an increased audiovisual processing (i.e., an enhanced N1 amplitude) was found in NT but not in autistic children; in the second stage, the audiovisual incongruence could be detected in both NT and autistic children (i.e., occurrence of pMMN); in the third stage, a sustained negative amplitude was found in the autistic non-McGurk group but not in either the NT or the autistic McGurk group. Furthermore, we found that EEG signals could decode the audiovisual congruent and McGurk/non-McGurk trials in all three groups but with different decoding time courses.

Behaviorally, we found that autistic children showed reduced audiovisual speech integration when compared to the NT group.

It was coherent with previous findings, which also employed the McGurk effect to investigate audiovisual speech integration and found a reduced audiovisual speech integration (less percentages of McGurk effect) in autistic children compared to NT children (e.g., Bebko et al. 2014; Feng et al. 2021, 2023; Stevenson et al. 2014). Previous studies revealed that reduced audiovisual integration was correlated with decreased language/communication abilities in autistic individuals (Feldman et al. 2018;

Mongillo et al. 2008). Consequently, the altered audiovisual speech integration may hinder the development of language abilities and communication skills in autistic children. Future studies could explore whether language abilities and communication skills could be improved by interventions targeting audiovisual speech integration.

Based on our modified three-stage process, in the first stage, the NT group showed an increased N1 amplitude in McGurk trials than in audiovisual congruent trials. However, the two autistic groups showed similar N1 amplitudes between McGurk/non-McGurk trials and audiovisual congruent trials. That is, autistic children showed an absence of increased N1 amplitudes compared with NT children. Our finding in the NT group was consistent with the previous finding in Lange et al. (2013). The increase in N1 amplitude may reflect the lack of facilitation by lip movements during the processing of audiovisual incongruent stimuli, in contrast to the facilitative effect observed in audiovisual congruent stimuli (Lange et al. 2013). Thus, the increased N1 amplitude might be seen as an early response to audiovisual incongruency. The absence of increased N1 response to audiovisual incongruency could be seen as the neural signatures of autism in early multisensory processing, specifically in audiovisual speech processing. Moreover, this alteration in N1 response to audiovisual incongruency might hinder higher-order speech perception and even language development in autism, which helps explain language and communication difficulties in autism.

However, unlike our finding, Roa Romero et al. (2015) found a decreased N1 amplitude for McGurk trials compared to audiovisual congruent trials in NT adults. They interpreted that this

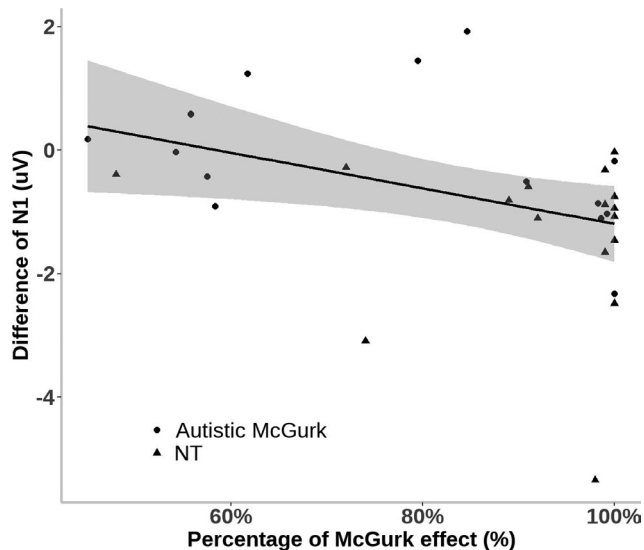


FIGURE 4 | Results of correlation analysis. Correlation between the McGurk effect and the difference of N1 amplitude (mean value in the 128–148 ms time window) across individual participants in the NT and autistic McGurk groups (Spearman's $\rho = -0.508$, $p = 0.004$).

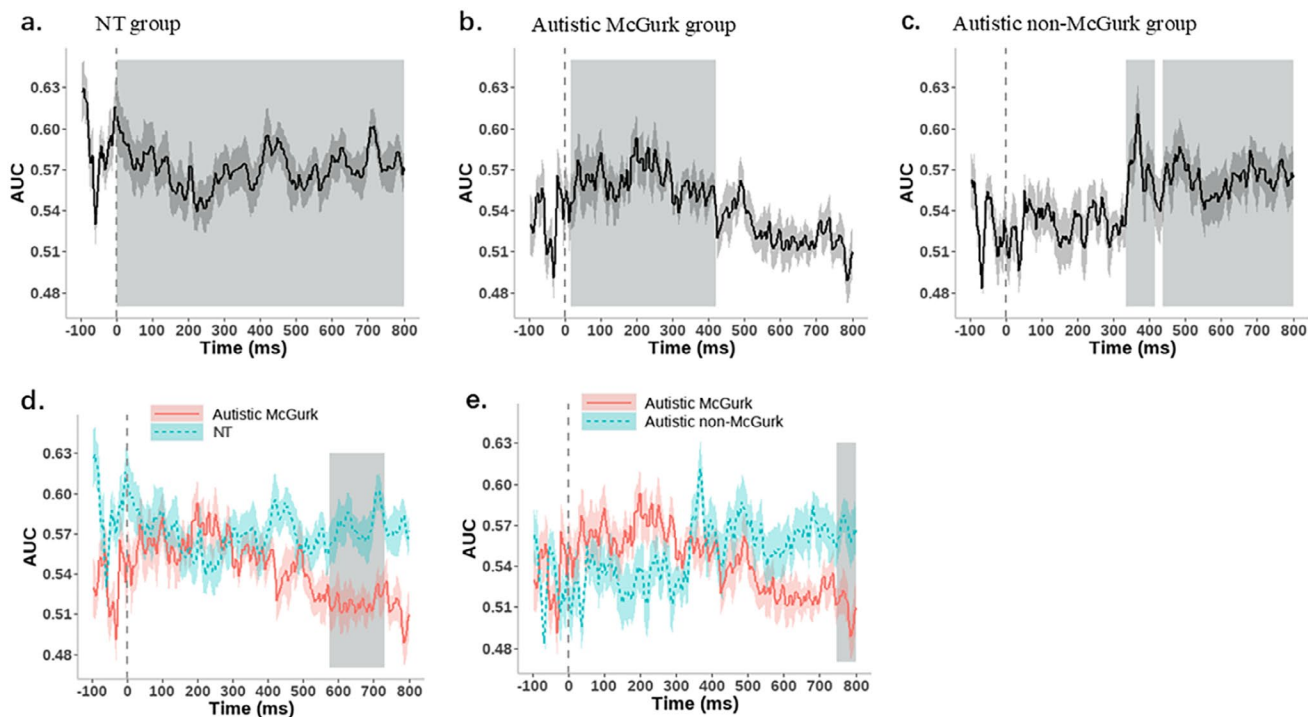


FIGURE 5 | MVPA results. Decoding accuracy showing the decoding performance across time between McGurk trials and audiovisual congruent trials for the NT group (a) and the autistic McGurk group (b), as well as between non-McGurk trials and audiovisual congruent trials in the autistic non-McGurk group (c). Decoding accuracy differences between NT and autistic McGurk groups (d) and between autistic McGurk and autistic non-McGurk groups (e). Significant time windows ($p < 0.05$, after cluster-based permutation) were indicated by gray areas.

decreased N1 amplitude possibly reflects early integrative processing. The discrepancy between our finding and that in Roa Romero et al. (2015) might be attributed to the age difference in participants. This suggests that the neural mechanisms underlying audiovisual speech integration differ between children and adults and are likely to develop with age. In adults, the decreased N1 likely reflects a stronger facilitation of lip movements for auditory processing, indicating that audiovisual integration begins at an early processing stage (Roa Romero et al. 2015). In children, however, the increased N1 likely represents a response to audiovisual incongruency. This suggests that children can detect the audiovisual incongruency but may not yet initiate audiovisual integration in such an early stage. Thus, the decreased N1 appears to be specific to adults with mature neural systems. This possibility warrants further investigations. Future studies could also track the developmental trajectory of the neural mechanisms underlying audiovisual speech integration across the lifespan.

In the first stage, we also found that magnitudes of the N1 difference waves between McGurk and audiovisual congruent trials were correlated with the strength of audiovisual speech integration by combining the NT group and the autistic McGurk group. It indicates that the early response to audiovisual incongruency (i.e., the increased N1 response to McGurk stimuli) might consequently affect the later audiovisual speech integration. However, in the autistic McGurk group, although we observed enhanced N1 amplitudes in McGurk trials compared with audiovisual congruent trials, the difference did not reach significance. It's possible that the lower responses to audiovisual incongruency in the autistic McGurk group weaken their audiovisual speech integration, which was consistent with the behavioral results.

In the second stage, we found that all three groups detected the audiovisual incongruence manifested by the pMMN when processing McGurk stimuli. The pMMN, first identified by Proverbio et al. (2018), indicates the detection of audiovisual incongruence in the McGurk stimuli. In our study, all three groups exhibited pMMN, which indicated that all three groups could detect the incongruence between the auditory information and the visual information in the McGurk stimuli. This indicates that autistic children in the present study have developed the ability to detect the audiovisual incongruency. Future studies could explore whether this ability generalizes to all autistic children.

In the third stage, the NT group and the autistic McGurk group showed similar amplitudes between McGurk trials and audiovisual congruent trials, indicating successful resolution of audiovisual incongruence. However, the autistic non-McGurk group showed a sustained negative amplitude in non-McGurk trials relative to audiovisual congruent trials, reflecting unsuccessful resolution of audiovisual incongruence. The sustained negative amplitude could be taken as a marker of unsuccessful audiovisual speech integration, especially for those autistic children who could not utter syllables.

Moreover, as for the EEG signals decoding, we found that the EEG signals decoded the McGurk and audiovisual congruent trials in the entire time course in the NT group but only in the earlier time course in the autistic McGurk group. At the same time, EEG signals decoded the non-McGurk and audiovisual congruent

trials only in the later time course in the autistic non-McGurk group. That is, in the two McGurk groups (the NT group and the autistic McGurk group) but not the autistic non-McGurk group, EEG signals could decode the two kinds of trials in the earlier time course. This indicated that children's stimuli processing in the earlier stages might affect their later audiovisual speech integration. We further speculated that the detection of audiovisual incongruency in the earlier processing stages might be a necessity of successful audiovisual speech integration. The failure in detecting audiovisual incongruency in the early processing stages might be one of the neural mechanisms underlying audiovisual speech integration difficulties in some autistic children.

The present study has several limitations. First, the participants were exclusively boys, limiting the generalizability of our findings to all populations. Future research should include female participants to determine whether these results are also applicable for autistic girls. Second, the IQ difference among the three groups reached marginal significance, causing potential influence on the results. Notably, the autistic McGurk group had slightly lower IQs than the autistic non-McGurk group, suggesting that higher IQ does not necessarily enhance audiovisual speech integration. To further clarify this relationship, future studies should systematically control for IQ effects in audiovisual speech processing. Thirdly, in the present study, the age range of participants was wide and we did not consider how age affected the neural mechanism underlying audiovisual integration. Since both audiovisual speech integration and auditory evoked ERP components develop with age (Ponton et al. 2000; Wallace and Stevenson 2014; Wallace et al. 2020), the neural mechanism underlying audiovisual speech integration likely changes across the lifespan. Future studies should recruit children within a narrower age range to better explore the neural mechanism and track its developmental trajectory.

In summary, we investigated the neural mechanisms underlying audiovisual speech integration in autistic children employing EEG in the present study. We found an increased early-stage audiovisual processing in NT children but not in autistic children. This altered early-stage audiovisual processing probably contributes to the reduced audiovisual speech integration in autistic children. In addition, it might be seen as a neural signature of autism in audiovisual speech processing. The present study also provided neural signatures corresponding to each stage of the modified three-stage process proposition in audiovisual speech integration proposed by us, thereby advancing the theoretical understanding of the audiovisual speech integration in both NT and autistic children.

Acknowledgments

The authors are grateful to Meirong Sun, Yaixin Zhang, and Lu Guo for their generous assistance in completing the study.

Funding

This work was supported by the Scientific Research Fund of Hunan Provincial Education Department, China [25A0005]; National Natural Science Foundation of China [32271116; 32571244; T2192932]; Clinical Medicine Plus X—Young Scholars Project of Peking University, the Fundamental Research Funds for the Central Universities; STI2030-Major Project [2021ZD0202600]; the Youth Foundation for Humanities and Social Science Research of Ministry of Education of China [24XJC190002].

Ethics Statement

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

Consent

We obtained oral consent from all autistic children and written consent from their parents before the experiment.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

References

- Altieri, N., D. B. Pisoni, and J. T. Townsend. 2011. "Some Behavioral and Neurobiological Constraints on Theories of Audiovisual Speech Integration: A Review and Suggestions for New Directions." *Seeing and Perceiving* 24, no. 6: 513–539. <https://doi.org/10.3389/fpsyg.2014.00257>.
- American Psychiatric Association. 2013. *Diagnostic and Statistical Manual of Mental Disorders*. 5th ed. American Psychiatric Press.
- Auyeung, B., S. Baron-Cohen, S. Wheelwright, and C. Allison. 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>.
- Bebko, J. M., J. H. Schroeder, and J. A. Weiss. 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>.
- Debnath, R., G. A. Buzzell, S. Morales, M. E. Bowers, S. C. Leach, and N. A. Fox. 2020. "The Maryland Analysis of Developmental EEG (MADE) Pipeline." *Psychophysiology* 57, no. 6: e13580. <https://doi.org/10.1111/psyp.13580>.
- Delorme, A., and S. Makeig. 2004. "EEGLAB: An Open Source Toolbox for Analysis of Single-Trial EEG Dynamics Including Independent Component Analysis." *Journal of Neuroscience Methods* 134, no. 1: 9–21. <https://doi.org/10.1016/j.jneumeth.2003.10.009>.
- Fahrenfort, J. J., J. van Driel, S. van Gaal, and C. N. L. Olivers. 2018. "From ERPs to MVPA Using the Amsterdam Decoding and Modeling Toolbox (ADAM)." *Frontiers in Neuroscience* 12: 368. <https://doi.org/10.3389/fnins.2018.00368>.
- Faul, F., E. Erdfelder, A. Buchner, and A.-G. Lang. 2009. "Statistical Power Analyses Using G* Power 3.1: Tests for Correlation and Regression Analyses." *Behavior Research Methods* 41, no. 4: 1149–1160. <https://doi.org/10.3758/BRM.41.4.1149>.
- Feldman, J. I., K. Dunham, M. Cassidy, M. T. Wallace, Y. Liu, and T. G. Woynaroski. 2018. "Audiovisual Multisensory Integration in Individuals With Autism Spectrum Disorder: A Systematic Review and Meta-Analysis." *Neuroscience and Biobehavioral Reviews* 95: 220–234. <https://doi.org/10.1016/j.neubiorev.2018.09.020>.
- Feng, S., H. Lu, Q. Wang, et al. 2021. "Face-Viewing Patterns Predict Audiovisual Speech Integration in Autistic Children." *Autism Research* 14: 2592–2602. <https://doi.org/10.1002/aur.2598>.
- Feng, S., Q. Wang, Y. Hu, et al. 2023. "Increasing Audiovisual Speech Integration in Autism Through Enhanced Attention to Mouth." *Developmental Science* 26: e13348. <https://doi.org/10.1111/desc.13348>.
- Hothorn, T., F. Bretz, and P. Westfall. 2008. "Simultaneous Inference in General Parametric Models." *Biometrical Journal* 50, no. 3: 346–363. <https://doi.org/10.1002/bimj.200810425>.
- Kleiner, M., D. Brainard, and D. Pelli. 2007. "What's New in Psychtoolbox-3? Perception, 36, ECVF Abstract Supplement."
- Knowland, V. C. P., E. Mercure, A. Karmiloff-Smith, F. Dick, and M. S. C. Thomas. 2014. "Audio-Visual Speech Perception: A Developmental ERP Investigation." *Developmental Science* 17: 110–124. <https://doi.org/10.1111/desc.12098>.
- Kushnerenko, E., T. Teinonen, A. Volein, and G. Csibra. 2008. "Electrophysiological Evidence of Illusory Audiovisual Speech Percept in Human Infants." *Proceedings of the National Academy of Sciences of the United States of America* 105, no. 32: 11442–11445. <https://doi.org/10.1073/pnas.0804275105>.
- Lange, J., N. Christian, and A. Schnitzler. 2013. "Audio-Visual Congruency Alters Power and Coherence of Oscillatory Activity Within and Between Cortical Areas." *NeuroImage* 79: 111–120. <https://doi.org/10.1016/j.neuroimage.2013.04.064>.
- Leach, S. C., S. Morales, M. E. Bowers, et al. 2020. "Adjusting ADJUST: Optimizing the ADJUST Algorithm for Pediatric Data Using Geodesic Nets." *Psychophysiology* 57: e13566. <https://doi.org/10.1111/psyp.13566>.
- Maris, E., and R. Oostenveld. 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., and J. MacDonald. 1976. "Hearing Lips and Seeing Voices." *Nature* 264: 746–748. <https://doi.org/10.1023/A:1005592401947>.
- Mongillo, E. A., J. R. Irwin, D. H. Whalen, C. Klaiman, A. S. Carter, and R. T. Schultz. 2008. "Audiovisual Processing in Children With and Without Autism Spectrum Disorders." *Journal of Autism and Developmental Disorders* 38: 1349–1358. <https://doi.org/10.1007/s10803-007-0521-y>.
- Nolan, H., R. Whelan, and R. B. Reilly. 2010. "FASTER: Fully Automated Statistical Thresholding for EEG Artifact Rejection." *Journal of Neuroscience Methods* 192, no. 1: 152–162. <https://doi.org/10.1016/j.jneumeth.2010.07.015>.
- 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>.
- Ponton, C. W., J. J. Eggermont, B. Kwong, and M. Don. 2000. "Maturation of Human Central Auditory System Activity: Evidence From Multi-Channel Evoked Potentials." *Clinical Neurophysiology* 111, no. 2: 220–236. [https://doi.org/10.1016/S1388-2457\(99\)00236-9](https://doi.org/10.1016/S1388-2457(99)00236-9).
- Proverbio, A. M., G. Raso, and A. Zani. 2018. "Electrophysiological Indexes of Incongruent Audiovisual Phonemic Processing: Unraveling the McGurk Effect." *Neuroscience* 385: 215–226. <https://doi.org/10.1016/j.neuroscience.2018.06.021>.
- Roa Romero, Y., J. Keil, J. Balz, M. Niedeggen, J. Gallinat, and D. Senkowski. 2016. "Alpha-Band Oscillations Reflect Altered Multisensory Processing of the McGurk Illusion in Schizophrenia." *Frontiers in Human Neuroscience* 10: 41. <https://doi.org/10.3389/fnhum.2016.00041>.
- Roa Romero, Y., D. Senkowski, and J. Keil. 2015. "Early and Late Beta-Band Power Reflect Audiovisual Perception in the McGurk Illusion." *Journal of Neurophysiology* 113, no. 7: 2342–2350. <https://doi.org/10.1152/jn.00783.2014>.

- Saint-Amour, D., P. De Sanctis, S. Molholm, W. Ritter, and J. J. Foxe. 2007. "Seeing Voices: High-Density Electrical Mapping and Source-Analysis of the Multisensory Mismatch Negativity Evoked During the McGurk Illusion." *Neuropsychologia* 45, no. 3: 587–597. <https://doi.org/10.1016/j.neuropsychologia.2006.03.036>.
- Stekelenburg, J. J., and J. Vroomen. 2012. "Electrophysiological Evidence for a Multisensory Speech-Specific Mode of Perception." *Neuropsychologia* 50, no. 7: 1425–1431. <https://doi.org/10.1016/j.neuropsychologia.2012.02.027>.
- Stevenson, R. A., J. K. Siemann, B. C. Schneider, et al. 2014. "Multisensory Temporal Integration in Autism Spectrum Disorders." *Journal of Neuroscience* 34: 691–697. <https://doi.org/10.1523/JNEUROSCI.3615-13.2014>.
- Wallace, M. T., and R. A. Stevenson. 2014. "The Construct of the Multisensory Temporal Binding Window and its Dysregulation in Developmental Disabilities." *Neuropsychologia* 64: 105–123. <https://doi.org/10.1016/j.neuropsychologia.2014.08.005>.
- Wallace, M. T., T. G. Woyanorski, and R. A. Stevenson. 2020. "Multisensory Integration as a Window into Orderly and Disrupted Cognition and Communication." *Annual Review of Psychology* 71: 193–219. <https://doi.org/10.1146/annurev-psych010419-051112>.
- Wechsler, D. 2014. *Wechsler Preschool and Primary Scale of Intelligence—Fourth CN Edition (WPPSI-IV CN)*, edited by Y. Li and Z. J. King-May Company China.
- Wheeler, B., and M. Torchiano. 2016. "lmPerm: Permutation Tests for Linear Models. R Package Version 2.1.0." <https://CRAN.R-project.org/package=lmPerm>.