Atypical audio-visual neural synchrony and speech processing in early autism

Xiaoyue Wang, Sophie Bouton, Nada Kojovic, Anne-Lise Giraud, Marie Schaer

Journal of Neurodevelopmental Disorders

RESEARCH

Open Access

Atypical audio-visual neural synchrony and speech processing in early autism

Xiaoyue Wang[1,2,4*], Sophie Bouton[2], Nada Kojovic[3], Anne-Lise Giraud[1,2†] and Marie Schaer[3†]

Abstract

Background Children with Autism Spectrum disorder (ASD) often exhibit communication difficulties that may stem from basic auditory temporal integration impairment but also be aggravated by an audio-visual integration deficit, resulting in a lack of interest in face-to-face communication. This study addresses whether speech processing anomalies in young autistic children (mean age 3.09-year-old) are associated with alterations of audio-visual temporal integration.

Methods We used high-density electroencephalography (HD-EEG) and eye tracking to record brain activity and gaze patterns in 31 children with ASD (6 females) and 33 typically developing (TD) children (11 females), while they watched cartoon videos. Neural responses to temporal audio-visual stimuli were analyzed using Temporal Response Functions model and phase analyses for audiovisual temporal coordination.

Results The reconstructability of speech signals from auditory responses was reduced in children with ASD compared to TD, but despite more restricted gaze patterns in ASD it was similar for visual responses in both groups. Speech reception was most strongly affected when visual speech information was also present, an interference that was not seen in TD children. These differences were associated with a broader phase angle distribution (exceeding pi/2) in the EEG theta range in children with ASD, signaling reduced reliability of audio-visual temporal alignment.

Conclusion These findings show that speech processing anomalies in ASD do not stand alone and that they are associated already at a very early development stage with audio-visual imbalance with poor auditory response encoding and disrupted audio-visual temporal coordination.

Keywords Autism spectrum disorders (ASD), Gaze direction, Speech envelope, Visual motion, Audio-visual, Oscillation phase entrainment

[†]Anne-Lise Giraud and Marie Schaer are co-last authors.

*Correspondence:
Xiaoyue Wang
wangxiaoyuepdf@gmail.com
[1]Auditory Language Group, Department of Basic Neuroscience, University of Geneva, Geneva, Switzerland
[2]Institut Pasteur, Université Paris Cité, Hearing Institute, Paris, France
[3]Autism Brain & Behavior Lab, Department of Psychiatry, University of Geneva, Geneva, Switzerland
[4]Department of Medical Psychology and Ethics, School of Basic Medical Sciences, Cheeloo College of Medicine, Shandong University, Jinan, Shandong, China

Background

Newborns are immediately attracted to the human voice. In-utero exposure to speech sounds enables them to accurately discriminate speech sounds at birth [1–3]. Since vision develops with a delay relative to hearing, babies only progressively discover that vocal stimuli are related to facial movements. Unlike typically developing (TD) children, children with ASD do not show this primary interest in speech [4–7]. Instead, they tend to engage in slow and repetitive visual exploration of their environment, eventually leading to atypical interests [8–13]. This focus on visual aspects of their surroundings allows ASD children to explore the world at their own pace and avoid highly dynamic stimuli such as speech and biological motion [14–16], which are often perceived as overwhelming [17, 18].

A basic auditory dysfunction in ASD might lead to speech-processing anomalies that in turn cascade into a decreased interest in speech [19–24]. Atypical speech processing becomes apparent early in development, and the fact that the associated neural anomalies in delta, theta, and gamma oscillations, accurately predict the severity of future language deficits suggests that they are causal to difficulties in language comprehension and production [25–28]. The tendency of children with ASD to prefer static or slow visual processing [29–32] possibly exacerbates speech reception challenges by counteracting dynamic audio-visual interaction, a crucial process for speech reception in ecological (e.g., noisy) environments [33–36]. Accordingly, excessively long integration time windows for audio and visual stimuli have been reported in children with ASD [31, 32], implying disturbed audio-visual integration.

Two essential mechanisms participate in audio-visual integration. The first one is the relative timing of auditory and visual stimuli: when these stimuli fall about 250 ms apart, they are often perceived as a single event, potentially influencing each other (e.g. the McGurk effect [37, 38]). The second mechanism is the re-synchronization provoked by the stimulus in one sensory modality on the neural responses to the other one [39–44]. Orofacial visual movements typically precede speech onset and lead to an auditory re-synchronization that sharpens responses to speech [39, 45]. Independent from the integration/fusion of the exact visual and speech content, visual re-synchronization enhances speech processing by boosting the tracking of the speech's syllabic structure. While anomalies of audio-visual (AV) binding (AV vs. A + V in ERP study) which tranditionally studied over short time windows in ASD are well documented [31, 32, 46–48], how audio and visual signals dynamically synchronize, involving the rhythmic synchronization of auditory and visual signals over longer periods, such as during natural speech exchanges, or when watching movies, remain hypothetical. Dynamic synchronization reflects the capacity for rapid re-synchronization with zero lag across an extended time course.

Auditory and visual sensory processing both operate rhythmically [49–53]. Visual speech information is characterized by a dominant 2–7 Hz rhythm (theta band [54]) and these quasiperiodic visual cues influence speech perception by modulating auditory neuronal oscillations within the same theta range at about 5 Hz [55–60], corresponding to the typical AV integration temporal time around 250ms [39, 61–64]. The resetting of auditory neural oscillations triggered by visual input [65] rhythmically enhances auditory processing [66], a phenomenon that is already observable in typical children [67]. Despite the documented presence of auditory processing anomalies in ASD at around 3 years [24], we still do not know whether they are associated with dynamic audio-visual synchrony anomalies.

This study aims to fill this gap by investigating with high-density EEG the dynamics of auditory and visual processing in young children with and without ASD, aged 1.13 to 5.56 years old, under naturalistic audio-visual conditions, i.e., when children are watching a popular cartoon adapted to their age. The goal is to compare the quality of the neural encoding/decoding of dynamic auditory and visual stimuli and audio-visual temporal coordination across groups.

Methods

Participants

Participants were selected from the Geneva Autism Cohort, a longitudinal study that aims to better understand the developmental trajectories in young children with ASD. This cohort's protocol has been detailed in previous studies [24, 68, 69]. In this study, we used clinical and behavioral assessments, as well as the electroencephalogram (EEG) recorded simultaneously with eye-tracking while children were watching popular cartoon videos.

The sample comprised 31 children with ASD (6 females, mean age = 3.09 years, SD = 0.91, age range: 1.74–5.14) and 32 TD peers (11 females, mean age = 2.95 years, SD = 1.31, age range: 1.31–5.56). Selection criteria for all participants included: age below 6 years, data collected during the participant's initial visit (i.e. at autism diagnosis for the autistic group), clear and accurate markers associated with movie onset, usable raw data for four different movies, and focus on the screen throughout all recordings. The age difference between the two groups was not significant (Bayesian independent samples t-test, BF10 = 0.287).

Autism clinical diagnoses were meticulously confirmed using standardized tools: either the Autism Diagnosis Observation Schedule-Generic (ADOS-G) [70] or the

Autism Diagnosis Observation Schedule, Second Edition (ADOS-2) [71]. Recruitment of participants occurred through specialized clinical centers and community-wide announcements. For the TD group, exclusion criteria included any suspicion of atypical psychomotor development, a history of neurological or psychological disorder, or having a first-degree relative with an autism diagnosis. Table 1 summarizes the clinical characteristics of the ASD and TD samples.

Stimuli and procedure

To explore cortical processing of audio-visual stimuli, we employed a passive and naturalistic task suitable for young children. This task involved viewing an age-appropriate French cartoon "TROTRO" [72–75] (example: https://www.youtube.com/watch?v=jT9C9WCIQr8%26t=81s). The selection of "TROTRO" was based on its cognitive accessibility and appeal to the target age group, including kids with ASD. The main character's verbal interactions enabled isolation of speech and visual motion for brain response analysis. Participants watched four 2.5-minute episodes in a consistent order. Visual engagement was monitored using Tobii Studio, an embedded application of Tobii TX300 eye tracking system. The videos were displayed on a screen with dimensions of 1200 pixels in height (29°38' visual angle) and 1920 pixels in width (45°53') with a refresh rate of 60 Hz, optimized for children's viewing comfort, with participants seated approximately 60 cm from the screen (Fig. 1A).

Eye-tracking acquisition and analysis

Gaze data were collected using the Tobii TX300 eye-tracking system (https://www.tobiipro.com), which operates at a sampling rate of 300 Hz. The cartoon was displayed in a frame that provided a visual angle of 26°47'(height) × 45°53'(width). Calibration was performed using a child-friendly procedure integrated into the Tobii system. To maintain consistency and reliability in data quality, we ensured constant lighting conditions in the testing room throughout all sessions. Special consideration was given to the youngest participants, who were seated on their parent's lap when they felt it more comfortable, a strategy that effectively minimized head and body movements that could have interfered with accurate data collection. We used the Tobii IV-T Fixation filter [76] to extract fixation data, offering precise measures of visual attention and gaze patterns. Inspired by previous findings of altered gaze distribution in autistic children [13, 69], this study used retinal stimuli around the gaze-fixation point as the source of visual information for subsequent analyses.

Audio and visual stimuli

We edited the movie soundtrack using Audacity v.2.2.1 to isolate speech excerpts, removing background noise like birdsong and music. Speech envelope data were extracted using the absolute value of the analytic signal [77], downsampled to 1000 Hz, and filtered with a 40 Hz zero-phase Butterworth filter. Visual motion data were tied to participants' gaze, focusing on stimuli within an 8-degree diameter [78, 79] around the retinal fixation point (318 × 318 pixels) (Fig. 1A1 & A2). The region was converted to grayscale, and luminance differences between successive frames exceeding a threshold of 10 were averaged to represent visual motion [80]. Visual motion was upsampled to match the EEG sampling rate (1000 Hz). Speech envelopes and visual motion were aligned, providing individualized stimulus data based on participants' gaze patterns for further analysis.

Stimulus features analysis

In order to assess shared information between speech envelope and visual motion, in autism and TD groups, we calculated mutual information (MI) scores, a dynamic metric, expressed in bits, which quantifies the reduction in uncertainty of one variable when another is observed [77, 78]. We calculated MI using the *quickMI* function from the Neuroscience Information Theory Toolbox [79]. The parameters for this calculation were set to 4 bins, no delay, and a p-value threshold of 0.001 [79]. For generating the MI scores, we concatenated all kept excerpts in the same sequence across subjects, separately for each stimulus feature and each group (ASD and TD). This process was followed by a comparative analysis of MI values between groups.

We only included stimuli corresponding to time periods with usable EEG signals. This resulted in slight variations in the stimulus duration for the ASD and TD groups, which were controlled for. No significant disparities in MI scores emerged between the two groups ($t(1,$

Table 1 Participants' demographic information and group comparison of behavioral tests

		Group					
		ASD (*N* = 31, 6 females)		TD (*N* = 32, 11 females)		Bayesian independent samples t-test	
		Mean	SD	Mean	SD	BF10	error%
Age(in years)		3.092	0.913	2.947	1.308	0.287	0.011
	Range	1.74–5.154		1.31–5.56			
ADOS		7.742	1.879	1.031	0.177	1.283E25	4.105E-28
	Range	4–10		1–2			

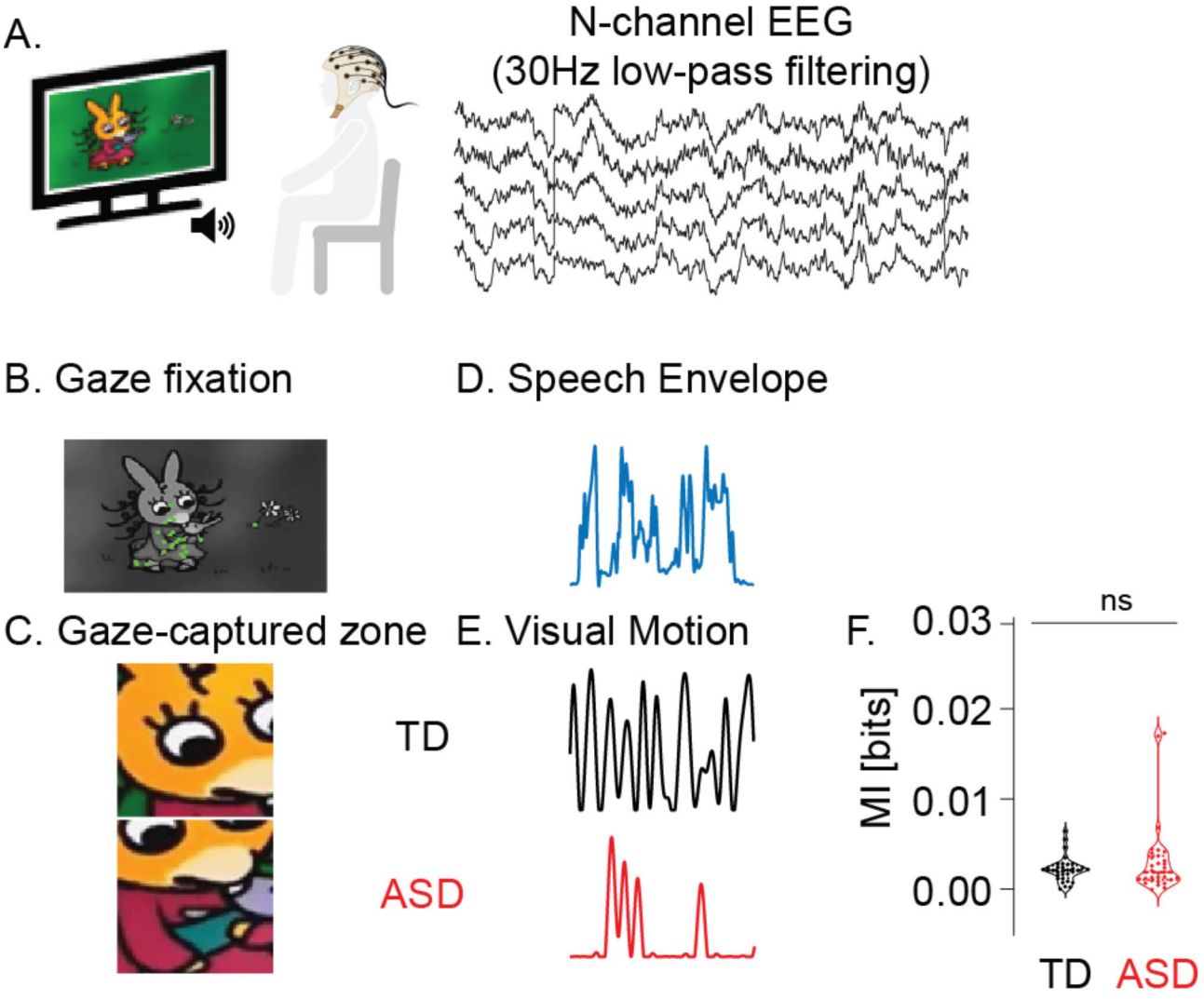

Fig. 1 Overview of Experimental Procedures and Features of Interest. **A** Experimental procedures **B** Gaze fixation. Example of individual gaze fixation points (green dots) on a black and white image; **C** Example of gaze-captured screen areas. Depiction of screen areas captured by the gaze of participants in ASD and TD (Typically Developing) groups. **D** Example of a stimulus speech envelope from the cartoon soundtrack. **E** Visual motion corresponds to the same stimulus in each group. **F** Comparison of speech envelope and visual motion. Mutual Information (MI) between ASD and TD groups (ns. $p > 0.05$)

61) = 1.250, $p = 0.216$, Cohen's d = 0.315, Fig. 1C), indicating that these minor differences did not lead to notable group differences in the shared information between the speech envelope and visual motion.

EEG acquisition and pre-processing

The EEG data were acquired using a 129-electrode (Hydrocel Geodesic Sensor Net (HCGSN) system (Electrical Geodesics, USA) at a sampling rate of 1000 Hz. During recording, the signals were subjected to real-time 0–100 Hz band-pass filtering. The reference electrode was positioned at the vertex (Cz). Data pre-processing was conducted using the EEGLAB v2019 toolbox within the MATLAB environment [80] and Cartool (https://sites.google.com/site/cartoolcommunity/). One hundred and ten channels were kept, excluding the cheek and neck electrodes to prevent contamination by muscle artifacts. EEG signals were filtered using a zero-phase fourth-order Butterworth bandpass (0.1–70 Hz) and a 50 Hz notch filter to eliminate power line noise. EEG data were visually inspected to remove movement artifact-contaminated periods. Bad channels were identified and excluded for exhibiting excessive signal amplitude. Eye blinks, saccades, electrical noise, and heartbeat artifacts were removed using independent component analysis (ICA). A spherical spline interpolation was used to interpolate the channels contaminated by noise using the ICA-corrected data. Finally, a common average reference was recalculated on the cleaned data, with an additional step of applying a 30 Hz low-pass filtering [81]. To ensure that all the EEG signals and stimulus features were on a similar scale and thus comparable, we normalized both the

EEG signals and stimulus features (i.e. speech envelope and visual motion) using the *nt_normcol* function (Noisetools: http://audition.ens.fr/adc/NoiseTools/*)*.

Temporal response functions (TRF)

To quantify how well EEG in ASD and TD children linearly varied with the stimulus features, we performed regularized regression (with ridge parameter λ) as implemented in the mTRF toolbox [82]. The TRF models the strength and direction of the brain's response to stimulus features, such as speech envelope or visual motion, at specific time lags. For the TRF modeling, we downsampled all signals to a rate of 100 Hz to accelerate computation.

Estimation of TRF using forward encoding models

We used a forward encoding model to predict EEG responses over time lags from 300 ms before to 300 ms after the stimulus. Separate univariate models were constructed for auditory (speech envelope, A-only) and visual (visual motion, V-only) stimuli. Additionally, a multivariate model (AV-joint) integrated both regressors, using trade-off weights to balance auditory and visual contributions, ensuring balanced representation of multisensory integration.

We compared the AV-joint model with A-only and V-only models to assess the benefits of audio-visual integration over unimodal processing. Using an n-fold leave-one-out cross-validation strategy, "generic" models were created to predict individual EEG data from TRFs derived from other participants. Model performance was optimized through a parameter search for the regularization parameter λ (see supplementary methods for λ selection). Pearson's correlation coefficient (r) between EEG signals with TRF-predicted signals was used to quantify model prediction accuracy per electrode. Then, correlations were averaged across participants to create a scalp-wide map of prediction accuracy. Finally, prediction accuracy was converted into Z-scores for statistical comparison by subtracting the surrogate data mean and dividing by its standard deviation. Surrogate distributions were generated by randomly shifting testing EEG segments, maintaining temporal structure. This analysis evaluated how accurately stimulus features were predicted from EEG data for each participant.

To quantitatively compare the accuracy between the ASD and TD groups, we used a cluster-based permutation test with 1000 randomization iterations, following the approach of Maris and Oostenveld [83]. Clusters were defined by considering both time and spatial electrode configurations, requiring each cluster to include at least two adjacent electrodes. A pivotal aspect of this approach was to ensure that the cluster-level type-I-error probability remained below the 0.05 threshold. This strategy was effective in controlling the family-wise error rate, maintaining it within the 5% type-I-error rate boundary.

Stimulus reconstruction using decoding models

We trained EEG decoders within a -300 to 300 ms post stimulus onset temporal window, using leave-one-out cross-validation and optimization to assess the accuracy of stimulus reconstruction (, i.e., speech envelope and visual motion). This approach allows us to identify the most informative segments for decoding, i.e. the time-lags with the highest EEG-stimulus synchronization. To determine the accuracy of stimulus reconstruction and select the best time lags, we used the Kruskal-Wallis test with Dunn's multiple comparisons test [84]. This non-parametric statistical method was chosen for its capability to handle variations in group means and variances across different conditions.

Low-frequency tracking of audio-visual signal

To explore whether the combined processing of auditory and visual stimuli relies on the tracking of audio-visual signals by low-frequency brain activity, we used a coherence-based and phase-based analytical framework [85]. This approach probed the interplay between neural responses and stimulus features by comparing their magnitude spectra and the phase relationship. Our analyses are centered on the delta and theta frequency bands, which are critical for effective integration of multimodal information [86].

Coherence analysis

We assessed individual responses to speech envelope and visual motion by computing magnitude-squared coherence for each trial and electrode using the *mscohere* function in Matlab, applying Welch's averaged modified periodogram method. The analysis spanned a frequency range from 0.1 to 30 Hz, in 0.33 Hz steps [87].

The analysis targeted delta (δ, ~4 Hz) and theta (θ, 4–8 Hz) frequency bands, identifying frequencies where coherence peaked most prominently for each stimulus condition. Statistical comparisons across groups and stimuli were conducted using the clusters identified through the method outlined in Sect. 4, combined with a nonparametric test and Dunn's multiple comparisons test [84]. Statistical significance was established using a surrogate-corrected coherence approach. Surrogate distributions were generated by randomly shifting the neural time course relative to the stimulus feature time courses, preserving their original temporal structure. This process was repeated 50 times for each stimulus condition to generate a robust surrogate distribution. The resulting coherence values were then standardized (Z-scored) against this distribution.

Phase analysis

We also performed a phase analysis by calculating the cross-power spectral density (CPSD) phase for each stimulus, electrode, and trial. This was done using the *cpsd* function in Matlab, employing parameters consistent with the coherence analyses. Phase values were determined based on the peak frequency identified in the coherence analysis. Group comparisons were conducted using Matlab's Circular Statistics Toolbox [88]. A two-way parametric ANOVA for circular data was performed to facilitate a nuanced comparison between pairs of conditions and groups, followed by post-hoc comparisons using the Watson-Williams multi-sample test [88]. Meanwhile, the Rayleigh test was used to investigate whether phase distribution was unimodal. Our focus was primarily on electrodes identified through TRF estimation outcomes.

Results

Atypical speech envelope processing in autistic children

We found distinct neural tracking patterns for the auditory and visual parts of the stimuli. Different scalp distribution patterns were observed between the two groups (Fig. 2) with cluster-correction $p < 0.05$. The ASD group had reduced neural response to the speech envelope relative to the TD group (Fig. 2A, top row). In contrast, there was no between-group difference in visual motion processing (Fig. 2A middle row).

A univariate stimulus reconstruction accuracy measure was compared between the ASD and TD groups, and aligned with the neural tracking findings, suggesting that speech processing is primarily impaired in autistic children. While reconstruction accuracy was comparable for speech envelope and visual motion ($p > 0.9999$, Fig. 2B) in the TD group, it was lower for speech than for visual motion ($p < 0.0001$, Fig. 2B) in the ASD group. These results suggest intact visual processing dynamic communicative stimuli in young children with ASD but atypical auditory processing.

Audiovisual integration anomaly in autism disrupts visual enhancement of auditory processing

We then explored whether speech anomalies in autism are limited to auditory processing difficulties or associated with audiovisual (AV) processing anomalies. First, the joint model suggested weaker AV representation in the ASD than in the TD group (Fig. 2A bottom row) and we found the expected stronger neural representation of the combined AV stimulus compared to individual single stimuli in both groups.

Distinct audiovisual integration patterns in ASD and TD children (Table 2; Fig. 3) were found by comparing

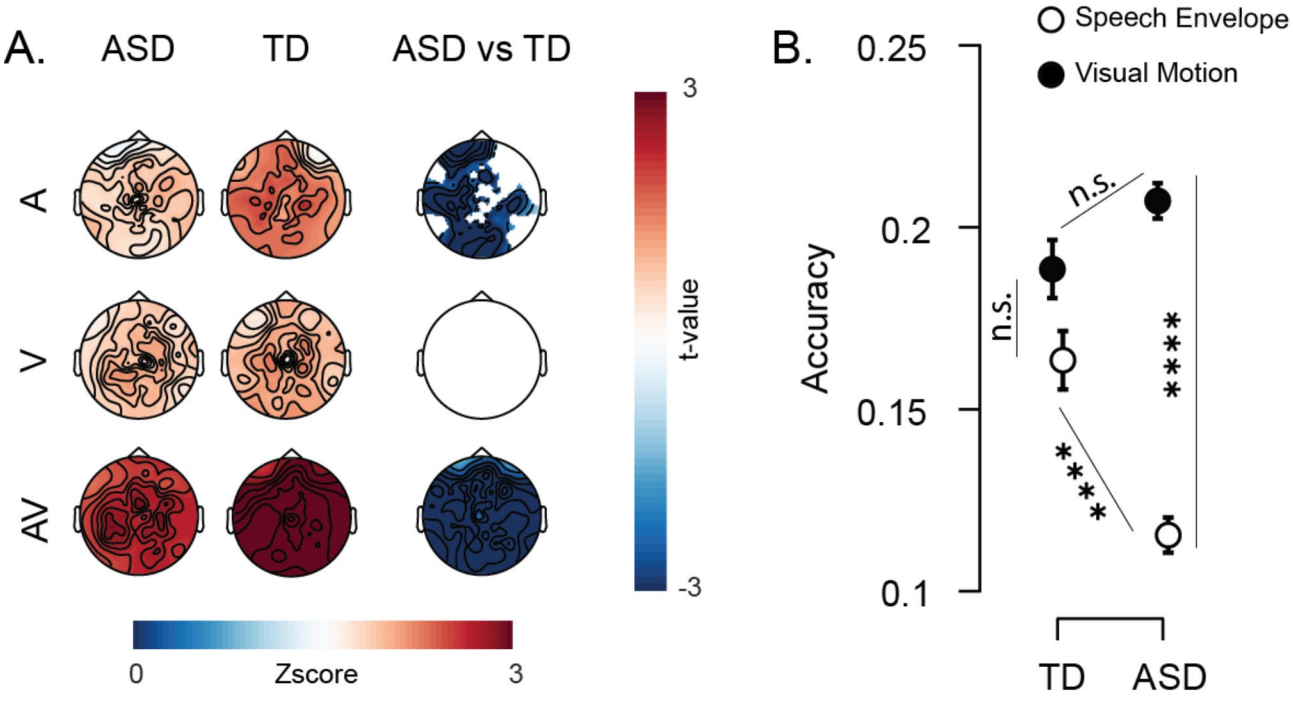

Fig. 2 Comparison of audio, visual, and AV models. **A**. Neural representations in ASD (left) and TD (middle) groups, for each model (A-speech envelope, V-visual motion, and AV joint) across all scalp electrodes. The right column shows EEG channels where significant group differences are observed using cluster-based nonparametric statistics ($p < 0.05$; with a positive t-value indicating greater predictability in the ASD group compared to the TD group). **B**. Stimulus reconstruction accuracy for speech envelope and visual motion in both groups. Error bars indicate the standard error of the mean. Significance levels are indicated as follows: 'ns' for $p > 0.05$ (not significant), * for $p < 0.05$, ** for $p < 0.01$, *** for $p < 0.001$, and **** for $p < 0.0001$

Table 2 The statistical difference among joint (AV) model and single models (A- &V-) for speech envelope and visual motion

	Dunn's multiple comparisons test	Mean rank diff.	Significant?	Adjusted *P* Value
ASD	speech envelope	-60.48	Yes	0.0304
	visual motion	-137.9	Yes	< 0.0001
TD	speech envelope	-51.09	No	0.1414
	visual motion	-58.47	Yes	0.0373

Table 3 The statistical difference of the decoding accuracy evolution for speech envelope and visual motion

	Dunn's multiple comparisons test	Mean rank diff.	Significant?	Adjusted *P* Value
Visual motion	ASD v.s.TD	-37.55	Yes	0.0003
Speech envelope	ASD v.s.TD	-13.32	No	0.8866
TD	V vs. A	-6.094	No	> 0.9999
ASD	V vs. A	-30.32	Yes	0.0065

A: speech envelope

V: visual motion

the decoding accuracies for speech envelope and visual motion across univariate (A-only and V-only) and multivariate (AV-joint) models. Notably, in the TD group, the accuracy of speech envelope reconstruction in the AV-joint model (concurrent speech envelope and visual motion) did not significantly differ from the A-only model ($p = 0.1414$, Fig. 3A). Conversely, in the ASD group, the speech envelope was less accurately decoded in the AV-joint model than in the A-only model ($p = 0.0304$, Fig. 3B), indicating a disruptive effect of AV integration on auditory processing specific to this group.

A decrease in visual motion reconstruction accuracy in the AV-joint model compared to the V-only model was observed in both groups (ASD group: $p < 0.0001$, TD group: $p = 0.0373$, Fig. 3), with a more pronounced decrement observed in the ASD group ($p = 0.0003$, Table 3). These findings suggest that integrating AV speech signals has a cost on on visual processing, and that this cost is higher in children with ASD.

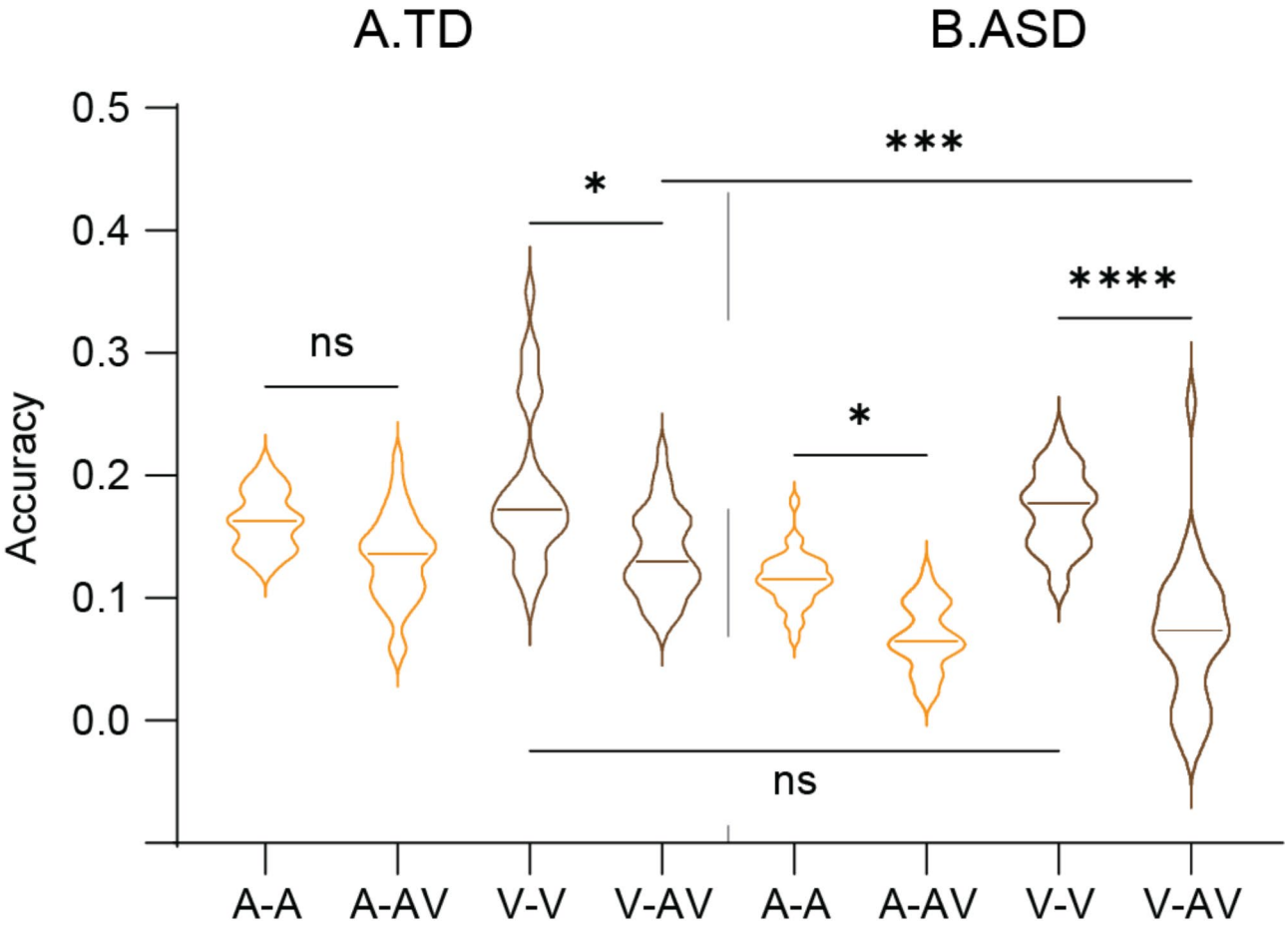

Fig. 3 Evaluation of decoding accuracy in TD **A** and ASD **B.** Stimulus reconstruction accuracy: speech envelope(A-) and visual motion(V-) in both the single-stimulus model (A-only = A-A and V-only = V-V) and the AV-joint model (A-AV, and V-AV)). Significance levels are indicated as follows: 'ns' for $p > 0.05$ (not significant), * for $p < 0.05$, ** for $p < 0.01$, *** for $p < 0.001$, **** for $p < 0.0001$. For additional details, see Supplemental Figs. 3 – 1

Surprisingly, we found distinct time-lags in auditory and visual decoding accuracies between ASD and TD when analyzing the temporal dynamics of audiovisual integration. In the TD group (Fig. 4A), auditory decoding reached significance at ~200 ms, while visual decoding took ~50 ms. In contrast, the ASD group showed the opposite pattern: ~200 ms for visual decoding and ~50 ms for auditory decoding (Fig. 4B). This reveals a visual lead in TD children, aligning with visual cues typically preceding sounds, but an auditory lead in autistic children, indicating a fundamental shift in sensory processing order.

Theta-range desynchronization of audio-visual responses in autism

To determine whether speech tracking anomalies in ASD stem from general stimulus/brain synchronization deficits or audiovisual integration issues, we analyzed the stimulus-response relationships in the delta (1–4 Hz) and theta (4–8 Hz) bands. Both groups showed higher stimulus/brain coherence in the delta than the theta band, with remarkably similar average coherence values and distribution patterns across groups (Fig. 5; Table 4). This indicates that the capacity of neural synchronization to auditory and visual stimuli is consistent in ASD and TD groups.

Given the preserved synchrony between brain activity and external stimuli in each modality, we then sought whether audio-visual integration anomalies, notably the inverted AV temporal patterns, are associated with a phase desynchronization of auditory and visual processing. In the delta band, we observed similar phase angles for both groups ($F(1,62) = 0.494$, $p < 0.470$), indicating comparable phase locking at this frequency, with small angles indicating the absence of delta band phase-shift between modalities. Yet, a significant group effect was observed in the theta band. In children with ASD the phase-shift amounted to 180 degrees and the group difference was significant ($F(1,62) = 12.05$, $p < 0.001$) (Fig. 6). The observed 180-degree phase shift in autism could suggest that auditory and visual information is out of sync: when one sensory modality is at its peak processing efficiency, the other is at its lowest, potentially leading to disjointed, even conflicting sensory processes.

AV phase-shift is related to auditory encoding accuracy in TD and visual encoding accuracy in autism

Finally, we explored the relationship between the AV phase-shift in the theta band and the accuracy of auditory and visual information reconstruction within an unimodal framework (Fig. 7). As expected, in TD children the AV phase-shift did not influence visual reconstruction accuracy ($r = 0.033$, $p = 0.858$), but there was a weak negative correlation between the phase-shift extent and speech reconstruction accuracy (r = -0.272, $p = 0.132$): when the AV phase-shift increased speech reconstruction accuracy decreased, which given the visual lead previously observed could suggest a causal effect. A different pattern was seen in children with ASD, with no relation between the phase shift extent and speech reconstruction

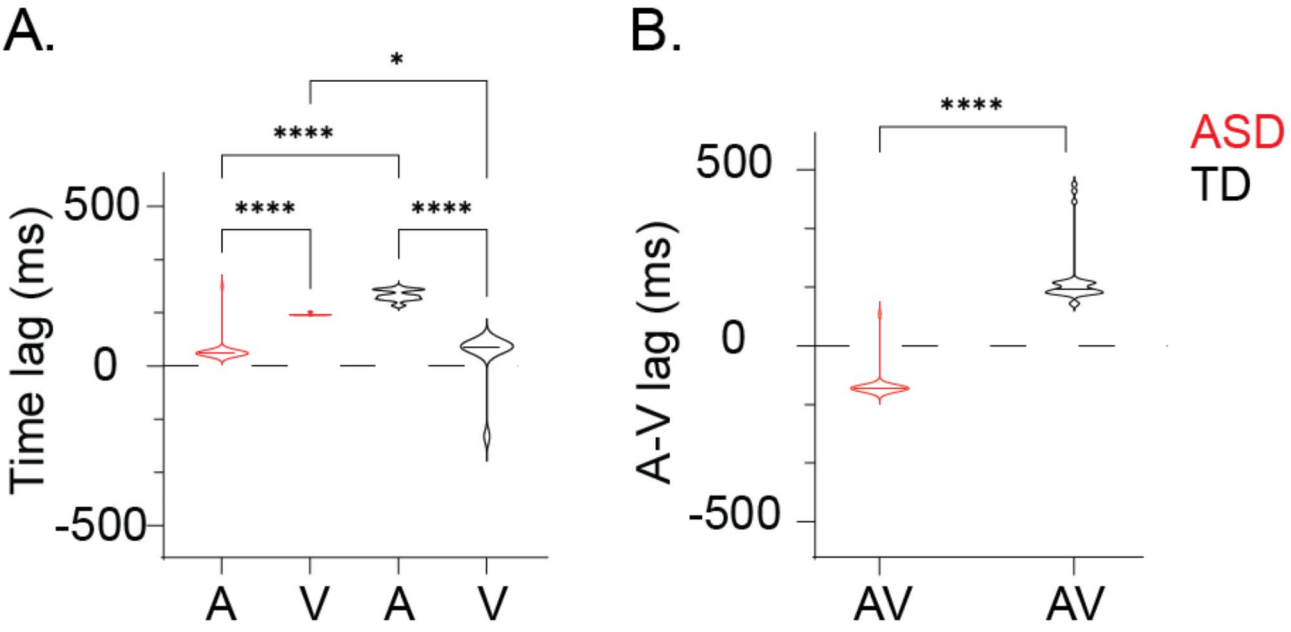

Fig. 4 Optimal EEG-stimuli time lag for ASD (red) and TD (black) groups. **A** depicts the optimal time-lag observed in the reconstruction of stimulus features in AV-joint model, specifically speech envelope (A-) and visual motion(V-); Positive values represent stimulus lead EEG signal. **B** illustrates the A-V time lag AV-joint model. Positive values represent V leads A. Significance levels are indicated as follows: ns > 0.05, *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, ****$p < 0.0001$

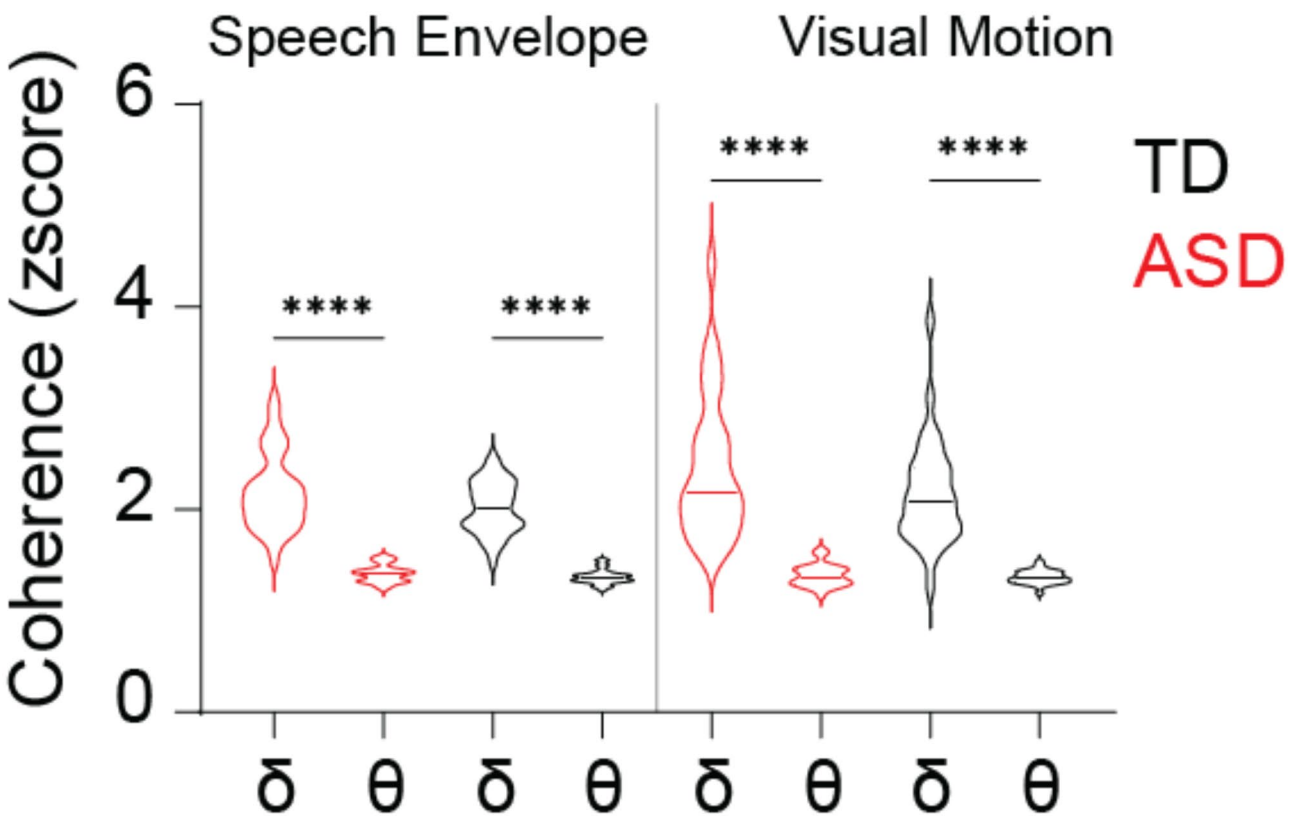

Fig. 5 Stimulus-response coherence in Theta and Delta Bands for ASD (red) and TD (black) groups. The plot displays the coherence between stimulus and response for Speech Envelope and Visual Motion. Error bars represent the standard error of the mean. The coherence levels are compared within the specific frequency bands of interest, highlighting potential group differences in sensory processing. Significance levels are indicated as follows: ns > 0.05, *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, ****$p < 0.0001$

Table 4 The statistical difference across groups and frequency bands in stimulus-response coherence

		Dunn's multiple comparisons test	Mean rank diff.	Significant?	Adjusted *P* Value
ASD	A	delta vs. theta	162.7	Yes	< 0.0001
ASD	V	delta vs. theta	191.5	Yes	< 0.0001
ASD	delta	A vs. V	-7.774	No	> 0.9999
ASD	theta	A vs. V	20.94	No	> 0.9999
TD	A	delta vs. theta	178.6	Yes	< 0.0001
TD	V	delta vs. theta	176.4	Yes	< 0.0001
TD	delta	A vs. V	-0.6563	No	> 0.9999
TD	theta	A vs. V	-2.844	No	> 0.9999
A	delta	ASD vs. TD	12.3	No	> 0.9999
A	theta	ASD vs. TD	28.19	No	> 0.9999
V	delta	ASD vs. TD	19.42	No	> 0.9999
V	theta	ASD vs. TD	4.406	No	> 0.9999

A: speech envelope

V: visual motion

accuracy ($r = 0.197$, $p = 0.288$; group*phase-shift t = -1.835, $p = 0.072$), but a weak negative correlation between the phase-shift extent and the accuracy of visual information reconstruction (r = -0.325, $p = 0.074$; group*phase-shift t = 1.062, $p = 0.293$), with larger AV phase-shifts linked to poorer visual reconstruction accuracy. Similarly, given the auditory lead observed in the group with ASD, this could suggest a causal effect.

Discussion

Using several analyses of the EEG recorded in very young children with and without ASD while they were watching a short animated movie, we confirmed previous results showing profound anomalies in the capacity to follow speech rhythms [24, 89], an essential prerequisite to speech comprehension. The present study goes beyond this observation by showing that children with ASD did not exhibit the natural dominance of auditory processing when exposed to natural audio-visual speech conditions. Instead, audio-visual processing was impacted by a temporal misalignment of these sensory inputs, which disrupted the predictive processing typically at play when perceiving speech.

Audio-visual integration anomalies interfere with sensory encoding in ASD

The synchronization of the two sensory modalities plays a pivotal role in understanding the communicative challenges observed in ASD. Our study reveals that speech

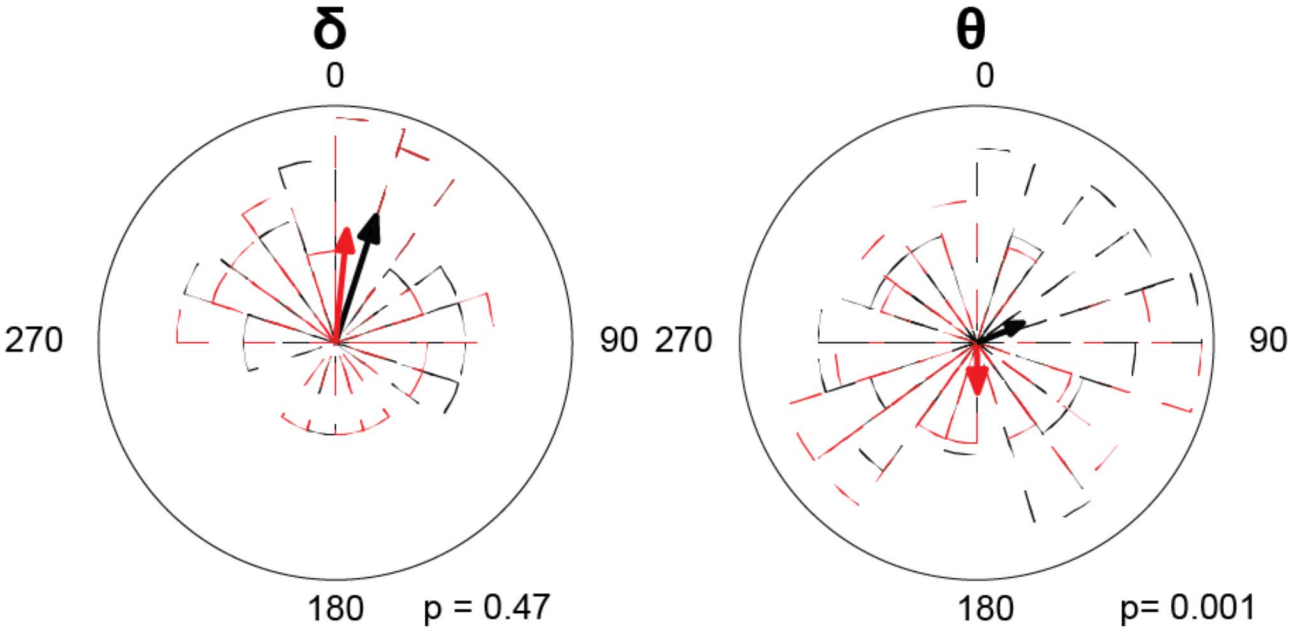

Fig. 6 Phase-shift distribution between speech envelope and visual motion. This figure shows the phase-shift distribution between the brain processes of speech envelope and visual motion stimuli for each group. The circular mean of the phase-shift across all subjects is indicated by colored lines: red for the ASD group and black for the TD group. Corresponding polar histograms in red (ASD) and black (TD) visually represent the distribution of phase-shifts for each group. Both groups were tested against the hypothetical uniform distribution of delta (rayleigh test, ASD: $p < 0.001$, rayleigh $r = 0.98$, TD: $p < 0.001$, rayleigh $r = 0.98$) and theta phase (rayleigh test, ASD: $p < 0.001$, rayleigh $r = 0.95$, TD: $p < 0.001$, rayleigh $r = 0.96$)

processing anomalies in ASD from an early developmental stage [19–24] are not merely isolated auditory deficits but are deeply connected to the integration of auditory and visual information, a process critical for effective communication, particularly in dynamic or complex listening environments [90, 91].

Our findings reveal a specific disruption in audio-visual integration among children with ASD, manifesting in visual dominance and temporal disorganization in auditory and visual processing. This disruption sharply contrasts with the expected auditory processing dominance [92] and might significantly contribute to the language development difficulties encountered by these children. In TD, the precedence of orofacial visual cues during speech facilitates auditory comprehension through predictive processing, optimizing the brain's synchronization to incoming speech signals [93]. In ASD, extended integration time windows and the lack of effective synchronization of auditory responses by visual signals (as evidenced by the atypical theta band phase-shifts) suggest they do not use visual cues to facilitate auditory speech processing, and that on the contrary auditory cues disrupt the visual processing in communicative situations.

Repercussions of disrupted audio-visual processing on speech tracking

Children with ASD exhibit visual motion tracking and processing capabilities comparable to their TD peers despite different scene analysis patterns, as previously observed [13, 94]. Within their preferred exploration zones, children with ASD process visual motion similarly to TD children [94] with an equivalent level of bottom-up excitability to visual stimuli [95, 96]. The univariate encoding results indeed suggest that the neural activity responsible for visual motion tracking operates similarly in both ASD and TD groups.

However, when visual processing co-occurs with speech processing, difficulties appear. Our multivariate modeling indicates that the neural encoding of audio-visual percepts in ASD children is less efficient, confirming that audiovisual contexts can disrupt brain responses to speech in this population [97]. Our study reinforces this crucial finding by showing that while autistic children encode single visual streams relatively well (visual motion tracking in univariate model), they struggle to concurrently encode auditory and visual streams (multivariate model).

Our study uncovers the potential underpinnings of the audiovisual integration difficulties observed in autism. The decoding results further indicate that while audio-visual integration interferes with visual processing in both groups, and that its impact on speech processing is particularly detrimental in the ASD group. Thus, the impairments in AV integration we observe are not merely additive but exacerbate sensory processing challenges in ASD. This framework explains that even though 12-month-old infants at risk for ASD explore faces and

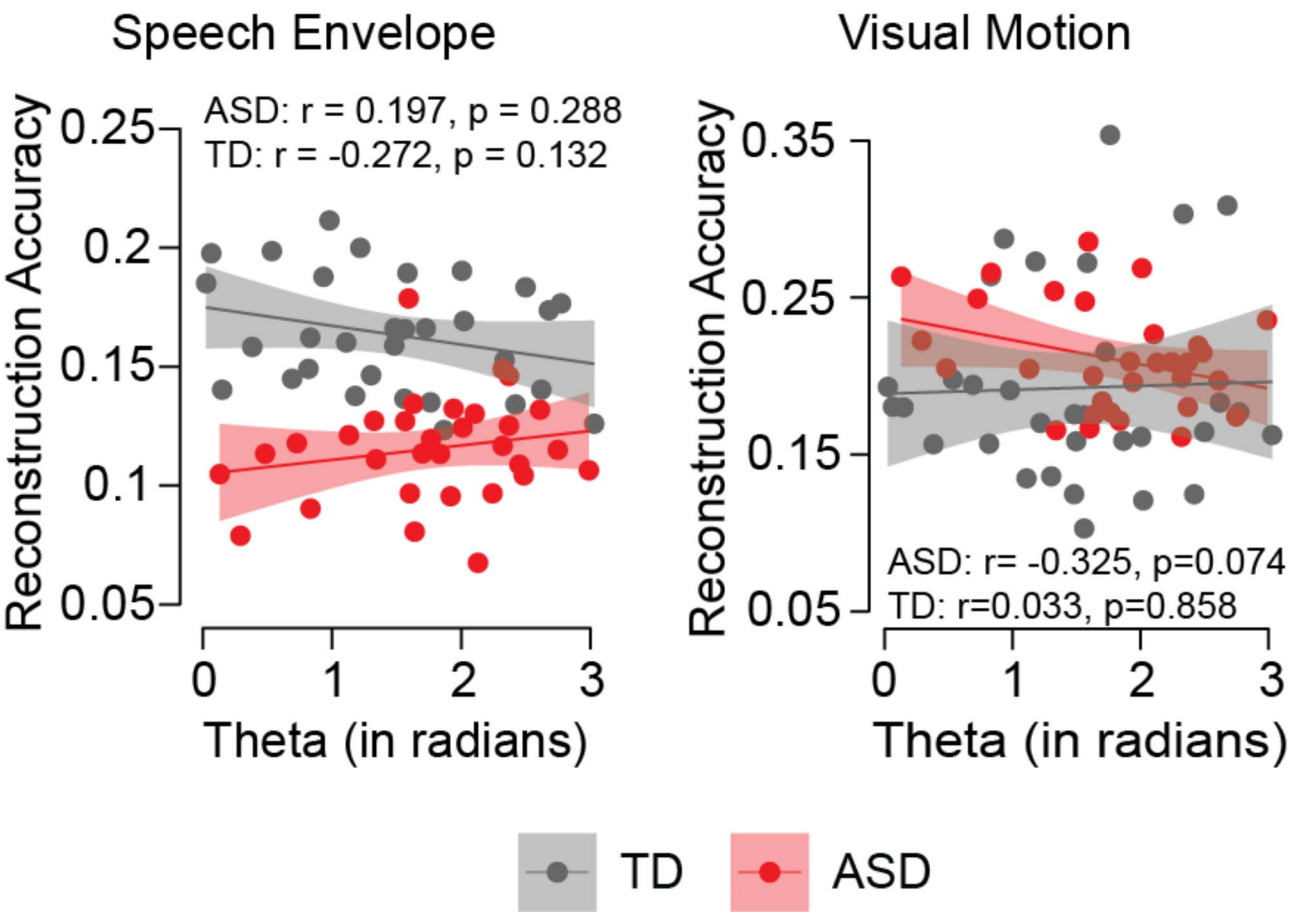

Fig. 7 The relationship between theta phase-shift and reconstruction accuracy of speech envelope (left) and visual motion (right) in ASD and TD. ASD group suggests a greater phase shift between speech envelope and visual motion positively correlates with speech reconstruction accuracy but negatively correlates with visual reconstruction, while reversely in TD group

mouths similarly to infants with no family history of autism [98], they cannot leverage audiovisual cues for language acquisition as do typical children.

Audio-visual temporal integration underlies speech impairment in autism

Audiovisual integration relies on the temporal alignment of sensory events, with visual cues enhancing auditory clarity, especially in noisy or ambiguous conditions [99–101]. Our findings confirm in TD children a visual lead (~ 50 ms) within a temporal window that is conducive to effective interaction and coordination between auditory and visual cues [102]. This window reasonably aligns with established models, positing a 200 ms integration period [39, 61–63], ranging from a 30 ms visual lag to a 170 ms of visual lead [61].

The precise timing of auditory and visual sequences is fundamental to audio-visual integration via predictive processing, whereby the brain leverages visual cues to anticipate and decode forthcoming auditory information. Here, phase-locking analyses in TD children show that the neural responses associated with auditory and visual processing exhibit a 90-degree phase shift, indicating that such a phase relationship optimizes a dynamic balance between the sensory streams, facilitating integration and enhancing perception and communication [101, 103]. The pivotal role of the theta frequency band in orchestrating audio-visual speech processing is robustly supported in the literature [56, 58, 59]. A pi/2 visual lead results in aligning visual information processing with the auditory inputs. This reliable phase alignment observed in TD children sharply contrasts with the broad phase distribution observed in the ASD group, signalling inconsistent audio-visual integration. In logic, reconstruction accuracy is a proxy of sensory encoding accuracy. Thus, in TD children, the relationship between phase-shift and reconstruction accuracy confirms the known reliance on visual cues to enhance auditory processing, with any misalignments adversely affecting speech information integration. Conversely, in ASD children, while speech encoding is weaker and overall less dependent on visual-auditory phase congruency, visual processing is vulnerable to strong AV resynchronization.

The atypical auditory lead (~50 ms) observed in ASD indicates that audio-visual integration is jeopardized and that the conventional sequence where visual information typically precedes auditory is inverted. Furthermore, the180-degree phase-shift in the neural activities associated with each stream reflects a profound disruption in temporal coordination, potentially leading to confusion or interpretation errors. Such a discrepancy underscores a critical deficiency in predictive processing in ASD, where, rather than synergistically enhancing each other, auditory and visual cues conflict, undermining the synthesis of coherent audio-visual perception [31, 32].

Our results are consistent with the notion that the phase of low-frequency neural oscillations is crucial for the temporal parsing in speech [104]. The anomaly in temporal encoding mechanisms described in our experiment is constrained by the temporal features provided by external stimulation to build a temporal reference frame. While delta oscillations have previously been linked to temporal predictability [102, 104], we observed here that sensory integration is affected by AV misalignment in the theta range, which is associated with atypical speech perception in ASD. AV integration primarily occurs at the syllable level with a typical tolerance to AV asynchrony around 250ms, which corresponds to the theta range [39, 61–64].

Conclusion

Our results show markedanomalies in audio-visual integration in young children with ASD that provide specific underpinnings for previous findings depicting disrupted speech rhythm tracking. They further reveal that disruption in audio-visual integration, manifesting as temporal desynchronization, impacts speech processing and contributes to the communicative challenges in autism. Our results also highlight the critical role of temporal processing in audio-visual integration and underscore the importance of characterizing these mechanisms in ASD. Moving forward, these insights could inform the development of targeted interventions aimed at regulating temporal speech processing and AV synchronization to improve communication in ASD children.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s11689-025-09593-w .

Supplementary Material 1

Supplementary Material 2

Acknowledgements
Not applicable.

Author contributions
Data acquisition and clinical resources: M.S., N. K.; Study design: X.W., N. K., M.S. and A-L.G.; Data analysis: X.W., S.B.; Writing-original draft: X.W.; Writing-review-editing: X.W., S.B., N. K., A-L.G. and M.S.; Supervision: A-L.G., and M.S.

Funding
This work is supported by grants from the Swiss National Science Foundation (#163859, #190084, #202235, & #212653 to M.S.), by the National Centres of Competence in Research (NCCR) Synapsy (Grant No. 51NF40–185897 to M.S.), by the Agence Nationale de la Recherche (ANR-21-CE28-0028 to S.B.), and by Evolving Language (Grant No. 51NF40_180888 to A-L.G). Additional support is provided by a grant from the Fondation pour l'Audition (FPA IDA11 to A-L.G), as well as by the Fondation Privée des Hôpitaux Universitaires de Genève (https://www.fondationhug.org) and the Fondation Pôle Autisme (https://www.pole-autisme.ch).

Data availability
The unprocessed datasets for this manuscript are not publicly available yet due to ongoing analysis as part of a longitudinal study. The results are expected to be published in the future. Once all data has been published, requests to access the datasets should be directed to Dr. Marie Schaer at marie.schaer@unige.ch. The custom MATLAB analysis scripts will be made available upon request to the xiaoyue.wang@pasteur.fr.

Declarations

Competing interests
The authors declare no competing interests.

Ethics approval and consent to participate
Informed consent was obtained from the parents of all participants prior to inclusion in the study. The research was conducted with the ethical standards set forth by the Ethics Committee of the Faculty of Medicine at the University of Geneva Hospital and adhered to the principles outlined in the Declaration of Helsinki.

Competing interests
The authors declare no conflict of interest.

Preprint servers
The manuscript was deposited as a preprint in bioRxiv with the license CC BY-NC-ND 4.0.

Received: 17 August 2024 / Accepted: 24 January 2025
Published online: 18 February 2025

References

1. Alegria J, Noirot E. Neonate orientation behaviour towards human voice. Int J Behav Dev. 1978;1:291–312.
2. Jusczyk PW, Bertoncini J. Viewing the development of speech perception as an innately guided learning process. Lang Speech. 1988;31:217–38.
3. Lecanuet J, Granier-Deferre C, Decasper A, Maugeais R, Andrieu A, Busnel M. Fetal perception and discrimination of speech stimuli; demonstration by cardiac reactivity; preliminary results. Comptes Rendus Académie Sci Sér III Sci Vie. 1987;305:161–4.
4. Williams JHG, Massaro DW, Peel NJ, Bosseler A, Suddendorf T. Visual–auditory integration during speech imitation in autism. Res Dev Disabil. 2004;25:559–75.
5. Iarocci G, Rombough A, Yager J, Weeks DJ, Chua R. Visual influences on speech perception in children with autism. Autism. 2010;14:305–20.
6. Feng S, Wang Q, Hu Y, Lu H, Li T, Song C, et al. Increasing audiovisual speech integration in autism through enhanced attention to mouth. Dev Sci. 2023;26:e13348.
7. Shic F, Wang Q, Macari SL, Chawarska K. The role of limited salience of speech in selective attention to faces in toddlers with autism spectrum disorders. J Child Psychol Psychiatry. 2020;61:459–69.
8. Klin A. Young autistic children's listening preferences in regard to speech: a possible characterization of the symptom of social withdrawal. J Autism Dev Disord. 1991;21:29–42.

9. Klin A, Lin DJ, Gorrindo P, Ramsay G, Jones W. Two-year-olds with autism orient to non-social contingencies rather than biological motion. Nature. 2009;459:257–61.
10. Klin A. Listening preferences in regard to speech in four children with developmental disabilities. J Child Psychol Psychiatry. 1992;33:763–9.
11. Dawson G, Meltzoff AN, Osterling J, Rinaldi J, Brown E. Children with autism fail to orient to naturally occurring social stimuli. J Autism Dev Disord. 1998;28:479–85.
12. Kuhl PK, Coffey-Corina S, Padden D, Dawson G. Links between social and linguistic processing of speech in preschool children with autism: behavioral and electrophysiological measures. Dev Sci. 2005;8:F1–12.
13. Kojovic N, Cekic S, Castañón SH, Franchini M, Sperdin HF, Sandini C et al. Unraveling the developmental dynamic of visual exploration of social interactions in autism. Büchel C, editor. eLife. 2024;13:e85623.
14. Federici A, Parma V, Vicovaro M, Radassao L, Casartelli L, Ronconi L. Anomalous perception of Biological Motion in Autism: a conceptual review and Meta-analysis. Sci Rep. 2020;10:4576.
15. Knight EJ, Krakowski AI, Freedman EG, Butler JS, Molholm S, Foxe JJ. Attentional influences on neural processing of biological motion in typically developing children and those on the autism spectrum. Mol Autism. 2022;13:33.
16. Todorova GK, Hatton REM, Pollick FE. Biological motion perception in autism spectrum disorder: a meta-analysis. Mol Autism. 2019;10:49.
17. Marco EJ, Hinkley LB, Hill SS, Nagarajan SS. Sensory processing in autism: a review of neurophysiologic findings. Pediatr Res. 2011;69:R48–54.
18. Todorova GK, Pollick FE, Muckli L. Special treatment of prediction errors in autism spectrum disorder. Neuropsychologia. 2021;163:108070.
19. Collet L, Roge B, Descouens D, Moron P, Duverdy F, Urgell H. Objective auditory dysfunction in infantile autism. Lancet. 1993;342:923–4.
20. Haesen B, Boets B, Wagemans J. A review of behavioural and electrophysiological studies on auditory processing and speech perception in autism spectrum disorders. Res Autism Spectr Disord. 2011;5:701–14.
21. Edgar JC, Fisk Iv CL, Berman JI, Chudnovskaya D, Liu S, Pandey J, et al. Auditory encoding abnormalities in children with autism spectrum disorder suggest delayed development of auditory cortex. Mol Autism. 2015;6:69.
22. Foss-Feig JH, Schauder KB, Key AP, Wallace MT, Stone WL. Audition-specific temporal processing deficits associated with language function in children with autism spectrum disorder. Autism Res off J Int Soc Autism Res. 2017;10:1845–56.
23. Wang X, Wang S, Fan Y, Huang D, Zhang Y. Speech-specific categorical perception deficit in autism: an event-related potential study of lexical tone processing in Mandarin-speaking children. Sci Rep. 2017.
24. Wang X, Delgado J, Marchesotti S, Kojovic N, Sperdin HF, Rihs TA, et al. Speech reception in Young Children with Autism is selectively indexed by a neural oscillation coupling anomaly. J Neurosci. 2023;43:6779–95.
25. Benasich AA, Gou Z, Choudhury N, Harris KD. Early cognitive and language skills are linked to resting frontal gamma power across the first 3 years. Behav Brain Res. 2008;195:215–22.
26. Benitez-Burraco A, Murphy E. The Oscillopathic Nature of Language deficits in Autism: from genes to Language Evolution. Front Hum Neurosci. 2016;10:120.
27. Cermak CA, Arshinoff S, Ribeiro de Oliveira L, Tendera A, Beal DS, Brian J, et al. Brain and Language associations in Autism Spectrum disorder: a scoping review. J Autism Dev Disord. 2022;52:725–37.
28. Morrel J, Singapuri K, Landa RJ, Reetzke R. Neural correlates and predictors of speech and language development in infants at elevated likelihood for autism: a systematic review. 2023 [cited 2023 Aug 28]; Available from: https://www.proquest.com/docview/2851820899/abstract/57E7B032F53541B0PQ/1
29. Crosse MJ, Foxe JJ, Tarrit K, Freedman EG, Molholm S. Resolution of impaired multisensory processing in autism and the cost of switching sensory modality. Commun Biol. 2022;5:1–17.
30. Jao Keehn RJ, Sanchez SS, Stewart CR, Zhao W, Grenesko-Stevens EL, Keehn B, et al. Impaired downregulation of visual cortex during auditory processing is associated with autism symptomatology in children and adolescents with autism spectrum disorder. Autism Res off J Int Soc Autism Res. 2017;10:130–43.
31. Stevenson RA, Siemann JK, Schneider BC, Eberly HE, Woynaroski TG, Camarata SM, et al. Multisensory temporal integration in autism spectrum disorders. J Neurosci. 2014;34:691–7.
32. Stevenson RA, Segers M, Ferber S, Barense MD, Wallace MT. The impact of multisensory integration deficits on speech perception in children with autism spectrum disorders. Front Psychol. 2014;5:379.
33. Alm M, Behne DM, Wang Y, Eg R. Audio-visual identification of place of articulation and voicing in white and babble noisea). J Acoust Soc Am. 2009;126:377–87.
34. Bertels J, Niesen M, Destoky F, Coolen T, Vander Ghinst M, Wens V, et al. Neurodevelopmental oscillatory basis of speech processing in noise. Dev Cogn Neurosci. 2023;59:101181.
35. Fleming JT, Maddox RK, Shinn-Cunningham BG. Spatial alignment between faces and voices improves selective attention to audio-visual speech. J Acoust Soc Am. 2021;150:3085–100.
36. Yuan Y, Lleo Y, Daniel R, White A, Oh Y. The Impact of Temporally Coherent Visual Cues on Speech Perception in Complex Auditory Environments. Front Neurosci [Internet]. 2021 [cited 2024 Mar 11];15. Available from: https://www.frontiersin.org/journals/neuroscience/articles/https://doi.org/10.3389/fnins.2021.678029/full
37. Guiraud JA, Tomalski P, Kushnerenko E, Ribeiro H, Davies K, Charman T, et al. Atypical audiovisual speech integration in infants at risk for autism. PLoS ONE. 2012;7:e36428.
38. Lindborg A, Baart M, Stekelenburg JJ, Vroomen J, Andersen TS. Speech-specific audiovisual integration modulates induced theta-band oscillations. PLoS ONE. 2019;14:e0219744.
39. Munhall KG, Gribble P, Sacco L, Ward M. Temporal constraints on the McGurk effect. Percept Psychophys. 1996;58:351–62.
40. van Wassenhove V. Speech through ears and eyes: interfacing the senses with the supramodal brain. Front Psychol [Internet]. 2013 [cited 2023 Aug 28];4. Available from: https://doi.org/10.3389/fpsyg.2013.00388
41. Choi I, Lee JY, Lee SH. Bottom-up and top-down modulation of multisensory integration. Curr Opin Neurobiol. 2018;52:115–22.
42. van Wassenhove V, Grant KW, Poeppel D. Visual speech speeds up the neural processing of auditory speech. Proc Natl Acad Sci U A. 2005;102:1181–6.
43. Kushnerenko E, Teinonen T, Volein A, Csibra G. Electrophysiological evidence of illusory audiovisual speech percept in human infants. Proc Natl Acad Sci. 2008;105:11442–5.
44. Musacchia G, Schroeder CE. Neuronal mechanisms, response dynamics and perceptual functions of multisensory interactions in auditory cortex. Hear Res. 2009;258:72–9.
45. Murray MM, Wallace MT, editors. The neural bases of multisensory processes. Boca Raton: CRC; 2011.
46. Baum SH, Stevenson RA, Wallace MT. Behavioral, perceptual, and neural alterations in sensory and multisensory function in autism spectrum disorder. Prog Neurobiol. 2015;134:140–60.
47. Foss-Feig JH, Kwakye LD, Cascio CJ, Burnette CP, Kadivar H, Stone WL, et al. An extended multisensory temporal binding window in autism spectrum disorders. Exp Brain Res. 2010;203:381–9.
48. Kwakye LD, Foss-Feig JH, Cascio CJ, Stone WL, Wallace MT. Altered auditory and multisensory temporal processing in autism spectrum disorders. Front Integr Neurosci. 2011;4:129.
49. Gao M, Lim S, Chubykin AA. Visual familiarity induced 5 hz oscillations and improved orientation and direction selectivities in V1. J Neurosci. 2021.
50. Lakatos P, Shah AS, Knuth KH, Ulbert I, Karmos G, Schroeder CE. An Oscillatory Hierarchy Controlling neuronal excitability and stimulus Processing in the auditory cortex. J Neurophysiol. 2005;94:1904–11.
51. Romei V, Brodbeck V, Michel C, Amedi A, Pascual-Leone A, Thut G. Spontaneous fluctuations in posterior α-Band EEG activity reflect variability in excitability of human visual areas. Cereb Cortex. 2008;18:2010–8.
52. Leszczynski M, Schroeder CE. The Role of Neuronal Oscillations in Visual Active Sensing. Front Integr Neurosci [Internet]. 2019 [cited 2024 Mar 11];13. Available from: https://doi.org/10.3389/fnint.2019.00032
53. Poeppel D, Assaneo MF. Speech rhythms and their neural foundations. Nat Rev Neurosci. 2020;21:322–34.
54. Chandrasekaran C, Trubanova A, Stillittano S, Caplier A, Ghazanfar AA. The Natural Statistics of Audiovisual Speech. PLOS Comput Biol. 2009;5:e1000436.
55. Arnal LH, Wyart V, Giraud AL. Transitions in neural oscillations reflect prediction errors generated in audiovisual speech. Nat Neurosci. 2011;14:797–801.
56. Young AW, Frühholz S, Schweinberger SR. Face and Voice Perception: understanding commonalities and differences. Trends Cogn Sci. 2020;24:398–410.
57. Aller M, Økland HS, MacGregor LJ, Blank H, Davis MH. Differential Auditory and Visual Phase-Locking are observed during Audio-Visual Benefit and Silent Lip-Reading for Speech Perception. J Neurosci. 2022;42:6108–20.
58. Hagan CC, Woods W, Johnson S, Calder AJ, Green GGR, Young AW. MEG demonstrates a supra-additive response to facial and vocal emotion in the right superior temporal sulcus. Proc Natl Acad Sci. 2009;106:20010–5.

59. Hagan CC, Woods W, Johnson S, Green GGR, Young AW. Involvement of right STS in Audio-Visual Integration for Affective Speech demonstrated using MEG. PLoS ONE. 2013;8:e70648.
60. Plöchl M, Fiebelkorn I, Kastner S, Obleser J. Attentional sampling of visual and auditory objects is captured by theta-modulated neural activity. Eur J Neurosci. 2022;55:3067–82.
61. van Wassenhove V, Grant KW, Poeppel D. Temporal window of integration in auditory-visual speech perception. Neuropsychologia. 2007;45:598–607.
62. Stevenson RA, Wallace MT. Multisensory temporal integration: task and stimulus dependencies. Exp Brain Res. 2013;227:249–61.
63. Massaro DW, Cohen MM. The paradigm and the fuzzy logical model of perception are alive and well. J Exp Psychol Gen. 1993;122:115–24.
64. Guilleminot P, Graef C, Butters E, Reichenbach T. Audiotactile stimulation can improve Syllable discrimination through multisensory integration in the Theta frequency Band. J Cogn Neurosci. 2023;35:1760–72.
65. Power A, Mead N, Barnes L, Goswami U. Neural Entrainment to Rhythmically Presented Auditory, Visual, and Audio-Visual Speech in Children. Front Psychol [Internet]. 2012 [cited 2023 Mar 3];3. Available from: https://doi.org/10.3389/fpsyg.2012.00216
66. Schroeder CE, Lakatos P, Kajikawa Y, Partan S, Puce A. Neuronal oscillations and visual amplification of speech. Trends Cogn Sci. 2008;12:106–13.
67. Keshavarzi M, Mandke K, Macfarlane A, Parvez L, Gabrielczyk F, Wilson A, et al. Atypical delta-band phase consistency and atypical preferred phase in children with dyslexia during neural entrainment to rhythmic audio-visual speech. NeuroImage Clin. 2022;35:103054.
68. Franchini M, Wood de Wilde H, Glaser B, Gentaz E, Eliez S, Schaer M. Brief report: a preference for Biological Motion predicts a reduction in Symptom Severity 1 Year later in preschoolers with Autism Spectrum disorders. Front Psychiatry. 2016;7:143.
69. Franchini M, Zoller D, Gentaz E, Glaser B, Wood de Wilde H, Kojovic N, et al. Early adaptive functioning trajectories in Preschoolers with Autism Spectrum disorders. J Pediatr Psychol. 2018;43:800–13.
70. Lord C, Risi S, Lambrecht L, Cook EH, Leventhal BL, DiLavore PC, et al. The Autism Diagnostic Observation Schedule—Generic: a standard measure of social and communication deficits associated with the spectrum of autism. J Autism Dev Disord. 2000;30:205–23.
71. Lord C, Rutter M, DiLavore P, Risi S, Gotham K, Bishop S. Autism diagnostic observation schedule: ADOS-2. West Psychol. J Psychoeduc Assess. 2012;32:88–92.
72. Trotro. l'anniversaire de nana. Storimages; 2013.
73. Trotro part en vacance. Storimages; 2013.
74. Trotro. et La boite a secrets. Storimages; 2013.
75. Trotro es tres amoureux. Storimages; 2013.
76. Olsen A. The Tobii I-VT fixation filter. Tobii Technol. 2012;21:4–19.
77. Weineck K, Wen OX, Henry MJ. Neural synchronization is strongest to the spectral flux of slow music and depends on familiarity and beat salience. Jensen O, Shinn-Cunningham BG, Zoefel B, editors. eLife. 2022;11:e75515.
78. Cover TM, Thomas JA, Entropy. Relative Entropy, and Mutual Information. Elem Inf Theory [Internet]. John Wiley & Sons, Ltd; 2005 [cited 2024 Mar 11]. pp. 13–55. Available from: https://onlinelibrary.wiley.com/doi/abs/10.1002/047174882X.ch2
79. Timme NM, Lapish C. A Tutorial for Information Theory in Neuroscience. eNeuro [Internet]. 2018 [cited 2024 Mar 11];5. Available from: https://www.eneuro.org/content/5/3/ENEURO.0052-18.2018.
80. Delorme A, Makeig S. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. J Neurosci Methods. 2004;134:9–21.
81. Jessen S, Fiedler L, Münte TF, Obleser J. Quantifying the individual auditory and visual brain response in 7- month-old infants watching a brief cartoon movie. NeuroImage. 2019.
82. Crosse MJ, Di Liberto GM, Bednar A, Lalor EC. The multivariate temporal response function (mTRF) toolbox: a MATLAB Toolbox for relating neural signals to continuous stimuli. Front Hum Neurosci. 2016;10:604.
83. Maris E, Oostenveld R. Nonparametric statistical testing of EEG- and MEG-data. J Neurosci Methods. 2007;164:177–90.
84. Dinno A. Nonparametric pairwise multiple comparisons in Independent groups using Dunn's test. Stata J. 2015;15:292–300.
85. Nozaradan S, Peretz I, Missal M, Mouraux A. Tagging the neuronal entrainment to beat and meter. J Neurosci. 2011;31:10234–40.
86. Clouter A, Shapiro KL, Hanslmayr S. Theta Phase synchronization is the glue that binds human associative memory. Curr Biol. 2017;27:3143–e31486.
87. Pefkou M, Arnal LH, Fontolan L, Giraud AL. theta-Band and beta-Band neural activity reflects independent Syllable Tracking and Comprehension of Time-compressed Speech. J Neurosci. 2017;37:7930–8.
88. Berens P, CircStat:. A *MATLAB* Toolbox for Circular Statistics. J Stat Softw [Internet]. 2009 [cited 2023 Jun 22];31. Available from: http://www.jstatsoft.org/v31/i10/
89. Jochaut D, Lehongre K, Saitovitch A, Devauchelle AD, Olasagasti I, Chabane N, et al. Atypical coordination of cortical oscillations in response to speech in autism. Front Hum Neurosci. 2015;9:171.
90. Puschmann S, Daeglau M, Stropahl M, Mirkovic B, Rosemann S, Thiel CM, et al. Hearing-impaired listeners show increased audiovisual benefit when listening to speech in noise. NeuroImage. 2019;196:261–8.
91. Haider CL, Suess N, Hauswald A, Park H, Weisz N. Masking of the mouth area impairs reconstruction of acoustic speech features and higher-level segmentational features in the presence of a distractor speaker. NeuroImage. 2022;252:119044.
92. O'CONNOR N, Hermelin B. Sensory dominance: in autistic imbecile children and controls. Arch Gen Psychiatry. 1965;12:99–103.
93. Arnal LH, Morillon B, Kell CA, Giraud A-L. Dual neural routing of Visual Facilitation in Speech Processing. J Neurosci. 2009;29:13445–53.
94. Liu W, Li M, Yi L. Identifying children with autism spectrum disorder based on their face processing abnormality: a machine learning framework. Autism Res. 2016;9:888–98.
95. Sinnett S, Soto-Faraco S, Spence C. The co-occurrence of multisensory competition and facilitation. Acta Psychol (Amst). 2008;128:153–61.
96. Cuppini C, Ursino M, Magosso E, Rowland BA, Stein BE. An emergent model of multisensory integration in superior colliculus neurons. Front Integr Neurosci [Internet]. 2010 [cited 2024 Mar 11];4. Available from: https://doi.org/10.3389/fnint.2010.00006
97. Irwin J, Harwood V, Kleinman D, Baron A, Avery T, Turcios J, et al. Neural and Behavioral Differences in Speech Perception for Children with Autism Spectrum Disorders within an Audiovisual Context. J Speech Lang Hear Res. 2023;66:2390–403.
98. Zoefel B, Archer-Boyd A, Davis MH. Phase entrainment of brain oscillations causally modulates neural responses to Intelligible Speech. Curr Biol. 2018;28:401–8. e5.
99. Noesselt T, Rieger JW, Schoenfeld MA, Kanowski M, Hinrichs H, Heinze H-J, et al. Audiovisual temporal correspondence modulates Human Multisensory Superior temporal Sulcus Plus primary sensory cortices. J Neurosci. 2007;27:11431–41.
100. Stacey PC, Kitterick PT, Morris SD, Sumner CJ. The contribution of visual information to the perception of speech in noise with and without informative temporal fine structure. Hear Res. 2016;336:17–28.
101. Han S, Chen Y-C, Maurer D, Shore DI, Lewis TL, Stanley BM, et al. The development of audio–visual temporal precision precedes its rapid recalibration. Sci Rep. 2022;12:21591.
102. Herbst SK, Stefanics G, Obleser J. Endogenous modulation of delta phase by expectation–A replication of Stefanics et al., 2010. Cortex. 2022;149:226–45.
103. Henry MJ, Obleser J. Frequency modulation entrains slow neural oscillations and optimizes human listening behavior. Proc Natl Acad Sci. 2012;109:20095–100.
104. Giraud AL, Poeppel D. Cortical oscillations and speech processing: emerging computational principles and operations. Nat Neurosci. 2012;15:511–7.
105. Stefanics G, Hangya B, Hernádi I, Winkler I, Lakatos P, Ulbert I. Phase entrainment of Human Delta oscillations can mediate the effects of Expectation on reaction speed. J Neurosci. 2010;30:13578–85.

Publisher's note