

Reviewed Preprint

v1 • December 31, 2025

Not revised

Reviewed Preprint

v2 • August 5, 2026

Revised by authors

✉ For correspondence:

michel.godel@unige.ch

* contributed equally to this work

Competing interests: No competing interests declared

Funding: See [page 36](#)

Reviewing editor: Jean-Paul Noel, University of Minnesota, United States

© 2025, Godel et al. This article is distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use and redistribution provided that the original author and source are credited.

From Syllables to Words: EEG Evidence of Different Age Trajectories in Speech Tracking and Statistical Learning in Infants at High and Low Likelihood for Autism

Michel Godel^{1,2}✉, Ana Fló³, Lucas Benjamin^{3,4,5}, Ghislaine Dehaene-Lambertz^{3,*}, Marie Schaer^{1,*}

¹Department of Psychiatry, University of Geneva School of Medicine, Geneva, Switzerland • ²Division of Adult Psychiatry, Department of Psychiatry, University Hospitals of Geneva, Geneva, Switzerland • ³Cognitive Neuroimaging Unit, CNRS ERL 9003, INSERM U992, CEA, Université Paris-Saclay, NeuroSpin Center, Gif/Yvette, France • ⁴Département d'étude Cognitives, École Normale Supérieure, Paris, France • ⁵Aix Marseille Univ, INSERM, INS, Inst Neurosci syst, Marseille, France

eLife Assessment

This **important** study uses longitudinal EEG to chart how neural tracking of syllables and word-level statistical structure develops over the first two years of life in infants at high and low likelihood for autism and links these measures to verbal outcomes at 18-20 months. The strength of evidence is **convincing**: the prospective longitudinal design, careful data-quality handling, and partial least squares analyses are appropriate and well executed, though some interpretations of the group differences in syllable tracking, along with the possible contributions of multilingual exposure and sleep state during recording, warrant caution. The work will be of interest to developmental cognitive neuroscientists studying language acquisition and early neural markers of neurodevelopmental conditions.

<https://doi.org/10.7554/eLife.109901.2.sa3>

Abstract

Delayed onset of canonical babbling and first words is often reported in infants later diagnosed with autism spectrum disorder. Identifying the neural mechanisms underlying language acquisition in autism is therefore critical to inform early diagnosis, prognosis, and intervention strategies. In this study, we investigated two speech processing mechanisms previously identified as atypical in children and adults with autism: the neural ability to track syllables; and statistical learning, the capacity to detect speech regularities beneath surface variability. We recorded 83 longitudinal high-density EEGs from 44 infants (2.5–22.6 months) at high (HL) and low (LL) likelihood for autism and assessed their verbal outcomes at 20 months. Neural entrainment was measured at syllable and word frequencies during exposure to a multi-speaker stream of concatenated tri-syllabic words, followed by a word recognition test using ERP recording. Our findings revealed reduced tracking abilities at the syllabic level in HL infants, a measure that correlated with verbal outcomes. While HL infants did not exhibit deficits in statistical learning itself, they displayed reduced novelty orientation during the word recognition test, indicated by a reduced late ERP. By contrast, multi-talker variability temporarily disrupted word segmentation around 12 months in LL infants, but not in HL infants, potentially reflecting decreased sensitivity

to human voices variability in the HL group. These results emphasize the importance of longitudinal protocols employing online, implicit measures to track the hierarchical stages of speech processing in both HL and LL infants.

Introduction

Autism Spectrum Disorder (hereafter: autism) is a neurodevelopmental condition marked by early and pervasive challenges in social interaction and communication, coupled with repetitive behaviors and restricted interests (1). The prevalence of autism has risen over the past decades to an estimated 3.2% (2,3). It is frequently associated with language difficulties that greatly vary across individuals and age (4,5). Atypical babbling and reduced word recognition are among the earliest indicators of autism (6,7). Identifying and addressing these early difficulties is crucial to mitigating their long-term detrimental cascading effects on later verbal and non-verbal abilities (8–10). Achieving this, however, requires a deeper understanding of the neural mechanisms underlying language acquisition in infants who eventually develop autism compared to typically developing peers.

A key step in language acquisition is the ability to discover the deep linguistic structure that lies beneath a variable surface. In particular, a primary challenge is to perceive the chain of discrete words embedded in a continuous speech flow, a difficult problem for autistic individuals vividly described by Donna Williams: “the way my brain had broken down sentences into words left me with a strange and sometimes unintelligible message” (11). This capacity relies both on the correct identification of the successive phonemes in the spoken flow and their attribution to a given word, as, unlike written text, where spaces delineate words, spoken language lacks clear perceptual boundaries. Successful word segmentation relies on a complex and hierarchical integration of prosodic, phonetic, lexical, and contextual cues (12). Among these cues, Saffran and colleagues (1996) highlighted the role of statistical regularities in the speech stream, showing that after just two minutes of exposure to a stream of randomly concatenated four tri-syllabic non-words, 8-month-old infants could distinguish these words from syllable combinations spanning word boundaries (13). This ability was attributed to the learning of transition probabilities between adjacent syllables—the computation of the likelihood of syllable B coming after syllable A in the artificial speech, $P(B|A)$. In their experiment, where each word was followed by one of three other words, transition probabilities were 1.0 within a word and 0.33 across word boundaries. This ability has put forward the contribution of statistical learning to speech segmentation, notably in preverbal infants. It was later shown that this capacity is already available in sleeping neonates (14,15). Nowadays, statistical learning is recognized as a general learning mechanism, available at any age (16) and observed in a wide range of perceptual domains (17–19). It operates largely automatically, as it has been observed even in sleeping and comatose states (15,20). Importantly, a causal role of statistical learning in language acquisition difficulties has been proposed in several studies of developmental disorders (21–24).

Considering this background, infants who eventually develop autism may face challenges at two levels of auditory processing. First, they may struggle at a perceptual stage: following the rhythm of speech and robustly encoding syllables. Second, statistical learning itself might be impaired. Prospective longitudinal studies of infants at high likelihood for autism (HL) have become the standard methodology to explore how neural processing takes place before the emergence of a reliable diagnosis (25,26). HL status usually stems from a family history of autism and/or a condition strongly associated with autism, like some specific genetic syndromes (27).

Regarding the first level, several EEG studies have reported decreased cortical synchronization to speech syllabic modulations in both autistic children and infants at high likelihood for autism (28–31). Specifically, these studies identified a negative relationship between theta-range (4–7 Hz) power in response to naturalistic speech stimuli and verbal abilities in autistic children. This diminished synchronization may stem from an excitatory/inhibitory neuronal imbalance in auditory cortices (32–34), as well as from anomalies in early auditory perception. For instance,

studies have observed jittering or distorted frequency encoding in auditory brainstem responses in autism, which could disrupt subsequent processing stages, such as a correct identification of the phonemes (35–37).

A growing body of literature also emphasizes atypical statistical learning in individuals with autism, particularly when using linguistic stimuli (38). After repeated exposures to a syllable stream, 10 y.o. autistic children showed a magneto-encephalographic neural response that was not modulated by the statistical properties of the syllable sequences, indicating a deficit in statistical learning (39). Exposing autistic children and HL infants to similar stimuli, two independent fMRI studies found that the left temporo-parietal cortex, left amygdala, and basal ganglia were less activated by statistical cues (40,41). However, the difference between groups was mainly due to a lack of activation in autistic children and might be related to low-level sensory difficulties classically described in autism, amplified by the scanner noise rather than to a genuine lack of statistical learning capacities, as discussed by the authors themselves. Importantly, most studies exploring either cortical synchronization or statistical learning have focused on autistic children rather than HL infants, i.e., well beyond the age at which those mechanisms support language acquisition.

Building on the evidence linking statistical learning to early language development and its potential disruptions in autism, we sought to investigate the preverbal developmental trajectory of this ability in infants at low and high likelihood of autism. Using high-density EEG, we conducted a prospective longitudinal study to evaluate how infants process an artificial speech stream similar to the one used by Saffran et al.'s (1996), during their first two years of life. Our study had four primary goals: first, to map the infant developmental trajectory of auditory statistical learning, which remains incompletely understood even in typically developing infants (42); second, to identify the specific levels of word learning at which HL infants may show difficulties; third, to determine whether these difficulties were stable, worsened, or improved over time compared to LL infants, and fourth, to assess whether any of these early indicators could predict later verbal difficulties.

We recorded 83 high-density EEG sessions from 44 infants (19 LL, 25 HL) aged 2.5 to 22.6 months, using an experimental paradigm originally designed for sleeping newborns (15,43,44). During the experiment, infants were exposed to a stream composed of randomly concatenated syllables (random stream or RND) and to an artificial speech stream consisting of syllables of constant duration that formed tri-syllabic non-words (structure stream or STR) (Figure 1A). Following this learning phase, they listened to isolated triplets corresponding to words from the stream and to part-words, which spanned word boundaries. This design allowed us to obtain several neural measures, which should help identify the specific difficulties faced by autistic children. First, the regular presentation of syllables at a fixed duration elicits increased power and Phase Locking Values (PLV) at the syllable frequency (4 Hz), providing an efficiency measure of neural synchronization with the speech signal. Second, previous research has shown that if the regular word structure is detected, neural entrainment at the word frequency emerges, resulting in increased PLV at the word frequency (4 Hz / 3 syllables = 1.33 Hz). This measure reflects the ability to segment the stream into words through statistical computations (15,16,45,46). Finally, comparing ERPs to isolated words and part-words enables to assess subsequent word recognition through familiarity/surprise responses (15). For each electrophysiological measure, we assessed its association with participants' receptive and expressive verbal outcomes collected at 18–21 months of age.

To increase the cognitive demands of our task and elicit potentially larger differences between LL and HL participants, we introduced random speaker changes across syllables. This acoustic variability was intended to add a layer of complexity that may pose a particular challenge for autistic individuals to filter out (47,48). Despite this variability, LL participants were expected to disregard voice changes, as previous studies have shown that infants can normalize phonetic representations across voices (49), enabling them to focus on the transition probabilities between syllables and extract the statistical structure of the stream. Notably, statistical learning has been demonstrated in neonates under these conditions of voice variation (44). It is worth noting,

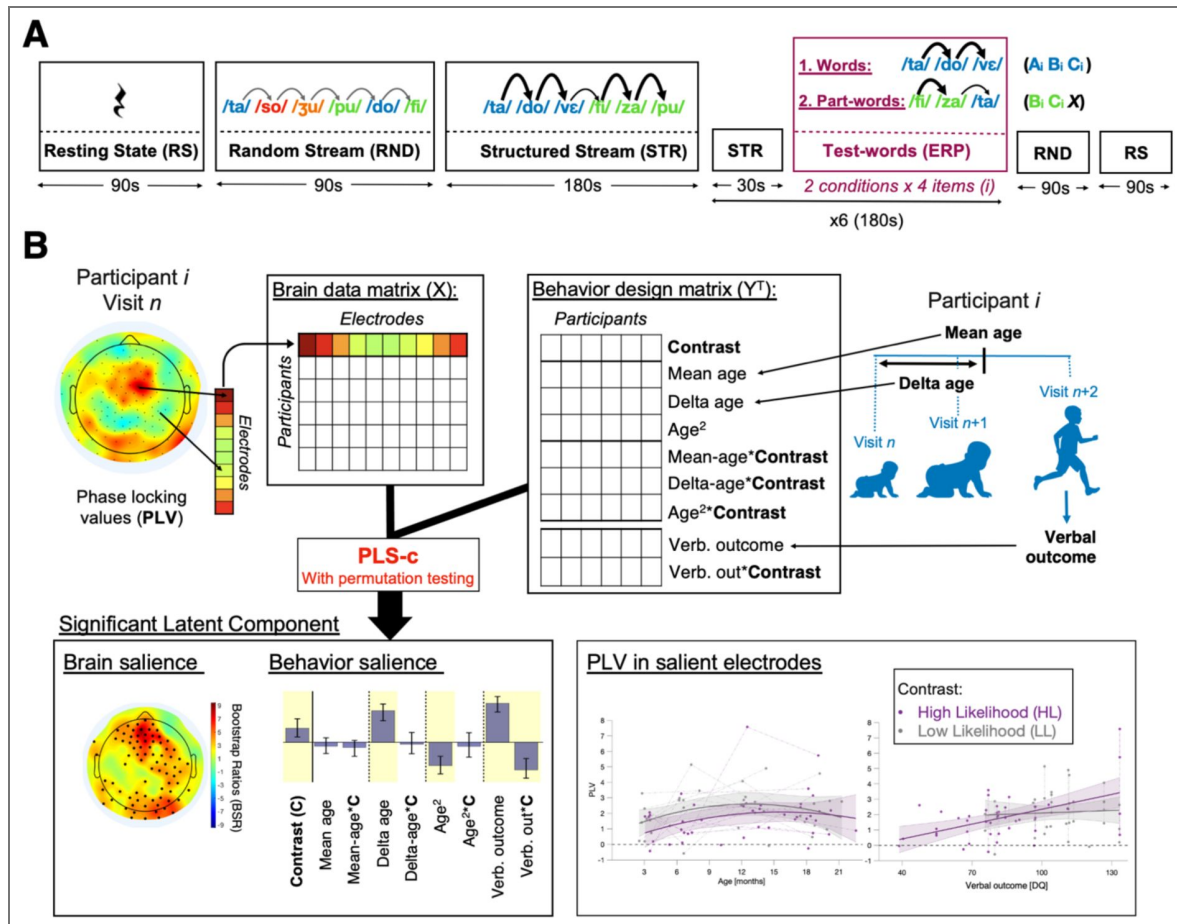


Figure 1.

A. Experimental procedure and multivariate statistical analyses. The learning part was sandwiched by a silent resting state (RS) and a random stream (RND) with even transition probabilities between syllables. This design accounted for the potential effect of time during the experiment and changes in vigilance state on neural entrainment measures. The learning segment consisted of a long structured (STR) stream where syllables were organized into four three-syllable words presented in random order with no repetition. Following this, six test-blocks were presented, each comprising 8 triplets from the words and part-words conditions with 2-second silences interleaved between items. To sustain learning, 30-second short STR streams were interspersed between test blocks. A 4.5s fade-in/out at the borders of each stream was included to minimize any perceptual anchor effect. The full procedure lasted ~17 minutes. Arrows' width schematically represents the transition probability magnitude. B. Pipeline for longitudinal partial least square correlation (PLS-c) analysis. Details are provided in the Methods section.

however, that our study was not designed to isolate or quantify the specific impact of speaker variability on statistical learning, as the experimental design did not include a baseline control condition omitting this acoustic variation.

To account for the multiple variables of the design, data were analyzed using longitudinal partial least square correlations (PLS-c, Figure 1B [↗](#)). We included participants' age, group (HL/LL), and verbal outcome to explain the EEG signal ([50,51](#)). Our analyses aim not only to provide a detailed characterization of syllable tracking and statistical learning in autism HL infants compared to typical infants, but also to track the evolution of these fundamental skills over the first two years of life in both groups of infants.

Results

Neural entrainment

Analyses of the entrainment at the syllabic rate (4 Hz) and the word rate (1.3 Hz) followed an identical logic. Phase-locking value (PLV) was employed as a measure of neural entrainment, as it directly quantifies neural synchronization to the stimuli. We examined whether neural entrainment was present in the different conditions across all participants while tracing the developmental trajectories of these responses across ages. To achieve this, we used partial least square correlation (PLS-c) ([51](#)). Briefly, PLS-c is a data-driven multivariate modelling approach designed to identify significant patterns of electrode clusters (from a brain data matrix containing electrophysiological measures, here PLV) and their associations with “behavioral” variables (from a behavioral design matrix, here age-related parameters). Patterns of brain x behavior associations are called latent components, and their statistical significance is evaluated using permutation testing ($n=1000$, Bonferroni correction for number of components tested, $\alpha=.006$). Brain and behavioral variables respective contributions to any significant latent component are tested with bootstrapping (500 random samples and replacement), with bootstrap ratios (BSR) greater than 2.3 indicating a stable contribution (for details, see the Materials and Methods section). The analysis aimed to confirm neural entrainment at the syllabic rate during both the random (RND) and structure (STR) streams and, more importantly, at the word rate during the STR stream only. Second, we investigated group differences (LL vs. HL) in entrainment by using a PLS-c with the entrainment at a given frequency as brain data and a behavioral design matrix including the group contrast, the age terms, and the verbal outcome variable. Finally, we assessed the dynamics of statistical learning by tracking changes in neural entrainment throughout the experiment.

Neural entrainment to syllables

The PLS-c analysis testing for syllabic entrainment in the RND and STR streams recovered one latent component ($p < .001$, $r = .75$, 93.1% explained covariance, Figure 2A-B [↗](#)). There was a significant contrast effect (entrainment at 4Hz vs. adjacent frequency bins, bootstrap ratio, BSR: 88.1), confirming syllable rate entrainment at the whole sample level. Interestingly, we observed a quadratic age effect (age² term) on the contrast, with a convex trajectory peaking around 12 months (BSR for contrast*age²: -4.0). Separate PLS-C analyses for each stream (i.e., RND and STR) yielded similar results, showing a comparable convex age trajectory in both streams (Figure 2 - figure supplement 1 [↗](#)). Because some infants were asleep during the recording session, particularly at younger ages, we performed a supplementary control analysis restricted to this sleeping subsample ($n = 25$ recordings, Figure 2 - figure supplement 2 [↗](#)). This PLS-c also identified a significant latent component ($p < .001$, $r = .78$, 85.1% explained covariance), with a significant contrast effect (BSR: 30.3) and a significant negative contrast*age² interaction (BSR: -2.5). These findings confirm that the convex age trajectory observed in the main analysis remains present and observable even in sleeping infants.

Regarding syllable entrainment differences between groups, we identified one significant latent component ($p < .001$; $r = .50$; 61.5% explained covariance, Figure 2C-D [↗](#)). This latent component was characterized by a significant effect of group (BSR: 3.1). LL infants showed overall stronger

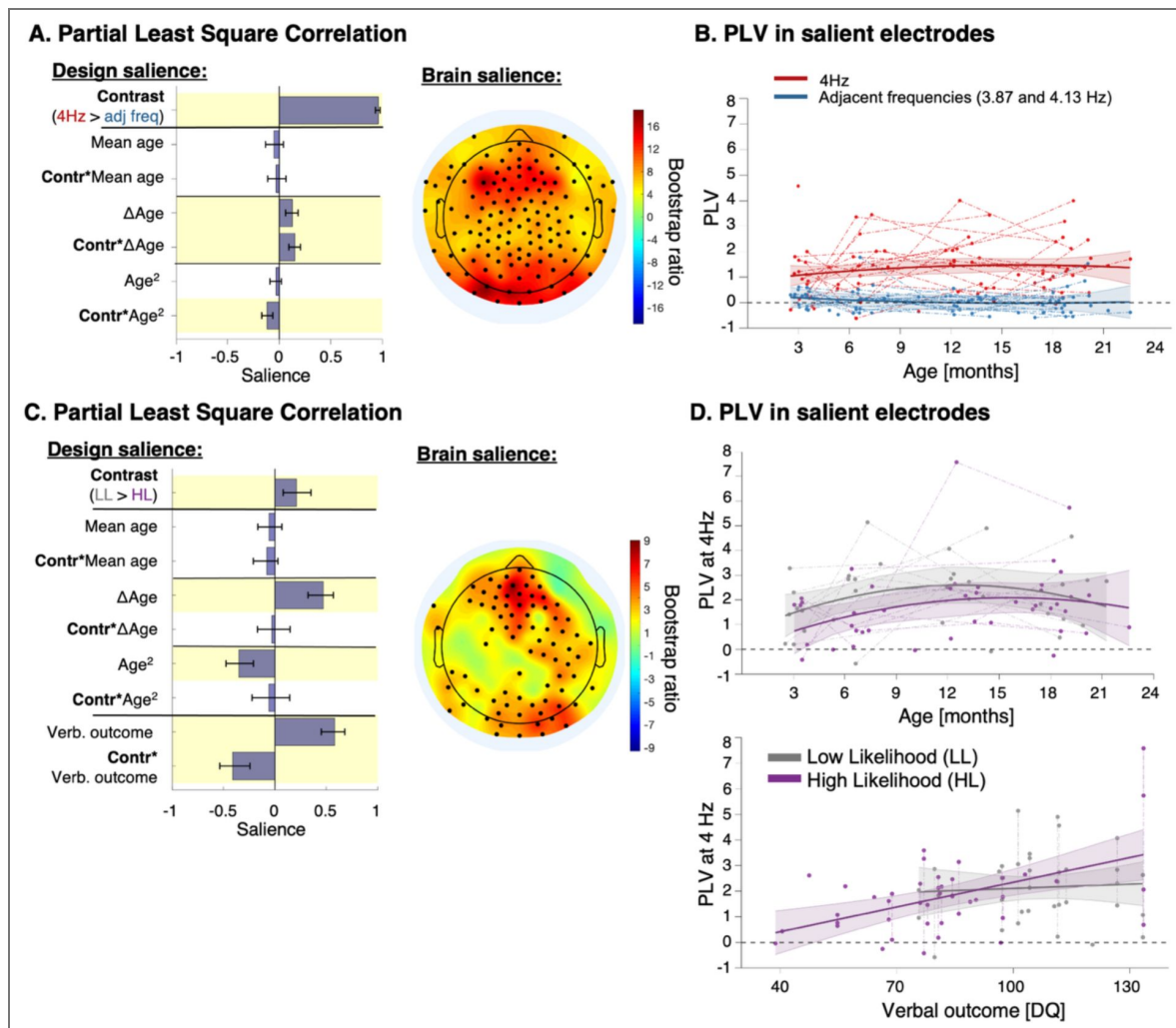


Figure 2. Neural entrainment to syllable rate (4 Hz): Main effect across all participants (Top row); Group differences (bottom row).

A. Design salience (left) and brain salience (right topography) derived from the significant latent component for neural entrainment to the syllable rate (4 Hz) using the targeted frequency (4 Hz) versus adjacent frequencies as a contrast. Significance (i.e., salience) was established through bootstrapping. Bars represent the mean of 500 random salience samples with replacement bootstrapping, and error bars indicate the 95% confidence interval. Yellow shading highlights variables that significantly contribute to the latent component, defined by a Bootstrap Ratio (BSR; mean of bootstrapping divided by standard deviation) > 2.3. The topography of the BSR values shows electrodes significantly contributing to the latent component (indicated by black dots, BSR > 2.3). **B.** Individual raw Phase Locking Values (PLVs) extracted from the salient electrodes identified by the latent component (black dots in the topography on 2A) are displayed. A fitted curve is included for visualization purposes only, produced by a mixed-effects model with a 95% confidence interval. This curve is intended solely to aid visualization, as the statistical relationships between EEG and behavioral variables are determined by the PLS-C analysis. **C.** PLS-C analysis of the differences between HL and LL groups is presented following the same format as in A. **D.** Individual raw PLVs extracted from the salient electrodes (black dots in subfigure 2C) are shown for visualization purposes only. At every age, a verbal developmental quotient (DQ) of 100 is expected in the general population.

syllable entrainment compared to HL participants. There was no significant age or age*group effect, indicating that syllable entrainment age-trajectories followed a similar convex shape in both groups. Moreover, the latent component comprised a positive effect of verbal outcome (BSR: 10.0), as well as a negative group*verbal outcome interaction effect (BSR: -5.9). This indicates that lower syllable entrainment was associated with poorer verbal performances at 18-21 months of age, predominantly in the HL group. To rule out the possibility that the association between syllable entrainment and verbal outcome was driven by concurrent measures taken at 18–21 months, we re-ran the PLS-c analysis excluding EEG data from the final visit (n=54 recordings kept). The resulting latent component remained significant ($p=.001$) and showed contributions from behavioral and EEG variables that were highly similar to those observed in the previous analysis, with a verbal outcome BSR of 7.1 and a group*verbal-outcome interaction BSR of -6.3 (Figure 2 -figure supplement 3 [3](#)).

Neural entrainment to words

The PLS-c on 1.33Hz entrainment vs. the entrainment at adjacent frequency bins during the STR stream resulted in one significant latent component ($p = .006$, $r = .33$, 30.7% explained covariance). There was a significant contrast effect (BSR: 29.6; Figure 3A-B [3](#)), confirming that at the whole sample level, infants' brains phased locked to the word rate and thus learned the regularities. This analysis also revealed a quadratic effect of age in the opposite direction of that observed for syllable entrainment, which showed a convex pattern (contrast $\times$ age² BSR: -4.0; Figure 2A-B [3](#)). By contrast, the age-related trajectory for word entrainment was concave, reaching a nadir at 12 months (contrast*age² BSR: 3.6). To confirm the specificity of the word effect, we performed a PLS-c analysis at 1.3 Hz during the RND condition, in which no entrainment was expected. This analysis didn't reveal any significant latent component. We further investigated word entrainment in sleeping participants (n=25 recordings), which yielded one significant latent component ($p<.001$, $r=.63$, 37.3% explained covariance, Figure 3 -figure supplement 1 [3](#)). Centro-frontal electrode contributed to this component, with a high contrast BSR (24.9), confirming a similar word entrainment pattern in the sleeping subsample. The contrast contrast*age² was also positive but not significant (1.1), suggesting a trend toward a U-shape age trajectory with a 12-month nadir in sleeping infants. Additional 18-21 month recordings would be required to confirm this trend.

Using group as the contrast variable (LL vs HL), we identified a significant latent component ($p=.002$; $r=.50$; 27.7% explained covariance) (Figure 3C-D [3](#)), but no reliable group effect (BSR<2.3). This indicates globally homogeneous neural tracking of words across infancy in both groups. There was also no effect of verbal outcome, nor significant interaction verbal outcome *group.

Given previous evidence of impaired statistical learning in autism, we conducted a supplementary PLS-c analysis restricted to the HL group to confirm that HL participants exhibited statistical learning abilities. Using the contrast (1.3Hz vs. adjacent frequencies) and the same age-related design variables, we confirmed significant word entrainment within the HL group (one significant latent component with $p=.002$, $r=.50$ and 38.2% explained covariance; contrast BSR=12.9, Figure 3 -figure supplement 2 [3](#)). Moreover, the salient electrodes contributing to the latent component showed a spatial distribution closely matching that observed in Figures 3A and 3C [3](#), confirming that the core neural mechanisms supporting statistical learning were consistent across groups. However, the significant age*group interaction, along with two additional effects: mean-age*group (BSR: 14.2) capturing cross-sectional age differences, and delta-age*group (BSR: -13.0), suggests distinct developmental trajectories between the two groups. Word entrainment in LL participants followed a more concave (U-shaped) age-trajectory than in HL participants (Figure 3D [3](#)). This interpretation was supported by a supplementary model comparison using Akaike Information Criterion (AIC) to evaluate constant, linear, and quadratic mixed-effects models on raw phase locking values (PLV) extracted from salient electrodes (black dots in Figure 3C [3](#)) within each group (52). For HL participants, the constant model had the lowest AIC (106.4) with a significant intercept above zero (.27 estimate, $p=.021$). In contrast, for LL participants, the quadratic model was best (AIC =93.4) with significant effects for the intercept (estimate 1.33, $p=.025$), age (estimate -.24, $p=.036$) and age² (.01 estimate, $p=.049$). A PLS-c analysis restricted to the HL group (contrast:

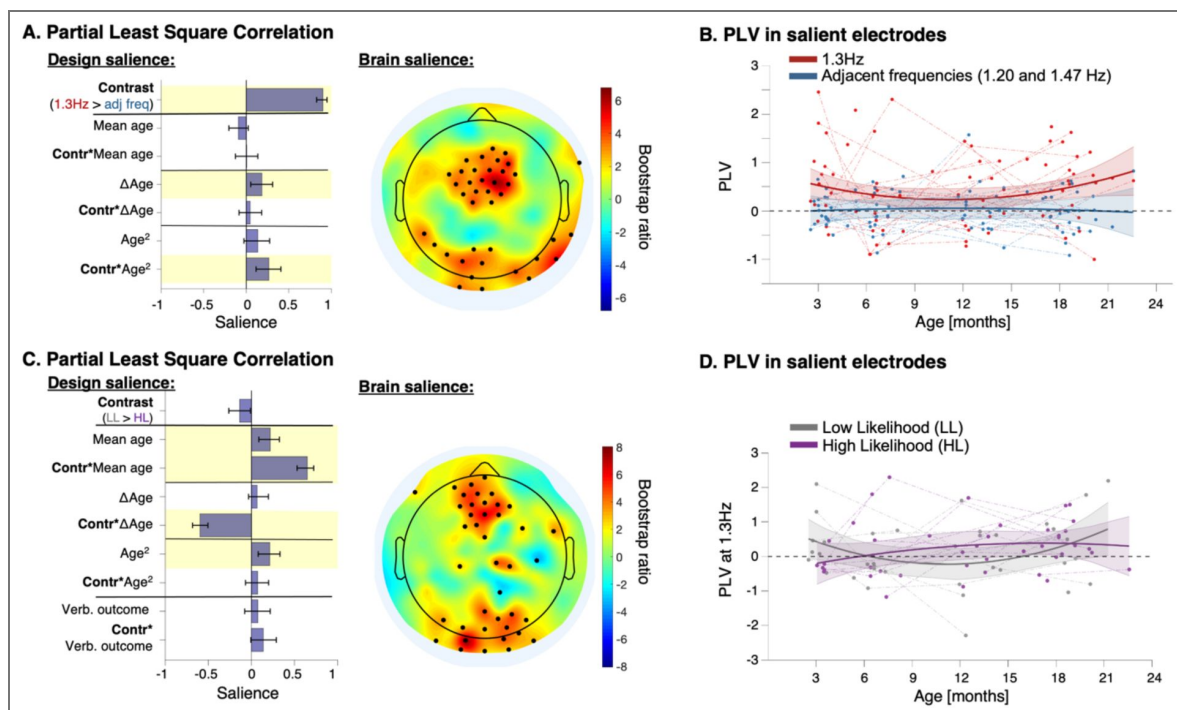


Figure 3. Neural entrainment to word rate (1.3Hz): Main effect across all participants

(Top row); Group differences (bottom row). **A.** Design salience (left) and brain salience (right) derived (through bootstrapping) from the significant latent component for neural entrainment to word rate. **B.** Individual raw PLVs extracted from the salient electrodes given by the latent component (black dots on subfigure A) with a fitted curve, for visualization purposes only. **C.** PLS-c applied on PLV at 1.3Hz with adjacent frequencies subtracted, using group as contrast, and verbal outcome (verbal developmental quotient [DQ] collected at 18-21 months) added as a design variable. **D.** Raw individual data extracted from the salient electrodes of the latent component (black dots on subfigure 2C) with a linear regression curve fitted for illustration purpose only.

1.33 Hz vs. adjacent frequency bins) further confirmed the absence of a U-shaped trajectory, showing no interaction with age² (BSR: <2.3, [Figure 3 -figure supplement 2](#)). In summary, statistical learning was reliably present and stable in HL participants, while LL participants showed a transient dip in performance, pointing to divergent developmental trajectories rather than differences in learning capacity.

Time course of the entrainment along the experiment

As a post-hoc analysis, we explored the temporal dynamics of statistical learning across the session using PLV data at each 1.5-second timeframe within a 120-second time window, averaged across the salient electrodes identified in the previous global analysis. Mixed-effects models fitted at each time frame comparing entrainment at 1.33Hz versus the adjacent frequency bins revealed that word-level entrainment emerged approximately 90 seconds after the onset of the structured stream (Supplementary material, [Figure S3](#)).

To gain further insight into the learning dynamics over time in the two groups, we conducted a PLS-c analysis comparing the syllabic and word entrainment time courses between LL and HL participants. HL/LL group was used as a contrast, while the brain data matrix (X) used PLV in both time and space dimensions, i.e., PLV in each electrode at each timeframe. For syllable entrainment, the analysis revealed a significant latent component ($p < .001$; see Supplementary Material S4). The group, age, and interaction BSRs were similar to those observed in the previous PLS-c presented in [Figure 2C](#). The group, age and verbal outcome parameters were mainly correlated (BSR>2.3) with the neural entrainment occurring ~90 seconds after the onset of the STR stream, coinciding with the time participants began tracking word boundaries (Supplementary material, [Figure S3](#)). This result suggests that the group differences in syllable entrainment, as shown in [Figure 2C-D](#), as their associations with verbal outcome are modulated by the structure of the stream (STR *versus* RND). We ran one additional PLS-c for each stream separately, using group as contrast. In both streams, the PLS-c yielded a significant latent component ($p < .001$ for RND and $p = .002$ for STR), with a positive group effect (BSR>2.3) in both latent components (Supplementary material, [Figure S5](#)). Most strikingly, the group*verbal outcome parameter reached significance exclusively within the STR latent component (BSR:-7.3). These results suggest that while syllable tracking is generally decreased in HL infants across both streams, its association with verbal outcome is prominently driven by the stream containing words (STR).

In contrast, a PLS-c analysis of the word-level entrainment time course revealed no significant latent components when comparing the two groups.

ERP analyses

Visual inspection of the grand-average ERP across all recordings showed an early response characterized by a frontal positivity accompanied by a posterior negativity, corresponding to the auditory response and developing over approximately the first second (word duration=750 ms). It was followed by a late component displaying a reversed spatial pattern from 1500 to 3000 ms (see Supplementary Material, [Figure S6](#)). ERP analyses using either condition or group as contrast variables were conducted in two temporal time windows: an early window sensitive to auditory processing [0-1000ms], and a later window [1500-3000ms] typically associated with higher-order cognitive responses, such as novelty detection and surprise-related activity ([53](#)).

Early ERP

We first tested whether a significant difference between word and part-word conditions was observed across all participants in the early [0-1000ms] time window (grand average topographies are displayed in Supplementary materials, [Figure S6](#)). We performed a longitudinal PLS-c contrasting part-word and word conditions. This analysis identified one significant latent component ($p < .001$; $R = .54$; 63.0% explained covariance, [Figure 4](#)). However, this component did not show any significant contribution from the condition contrast or its interaction with age variables (BSR<2.3). Instead, only the age-related variables (BSRs for age: -12.6; delta-age: -29.5; age²: 11.1) significantly contributed to the latent component. This indicates an age-related decline

in the amplitude of both the frontal positivity and the posterior negativity, following a slightly quadratic concave shape (Figure 4B). There was no evidence for any modulatory effect of the part-word/word condition on this developmental pattern. The same PLS-c, conducted separately in sleeping ($n=25$) and awake subsamples ($n=58$), yielded no significant latent component, indicating a consistent absence of early response to word novelty in both sleeping and awake infants.

A PLS-c analysis of the differential ERP to part-words vs. words using group as contrast (HL *versus* LL) with age-related and verbal outcome parameters did not reveal any significant latent component. This indicates that the absence of an early condition effect in the full sample (Figure 4) was not due to opposing response patterns between the LL and HL groups.

Late ERP

The PLS-C analysis in the late time window (1500 to 3000 ms), contrasting words vs. part-words, resulted in one significant latent component ($p<.001$, $R=.49$, 47.4% explained covariance) (Figure 5). The part-word vs. word contrast showed a significant contribution (BSR: 6.9), with a late frontal negativity and posterior positivity more pronounced in the part-word condition than in the word condition (see Supplementary materials, Figure S6 for grand average topographies). All age variables contributed significantly to the latent component (BSRs for age: -10.9; delta-age: -21.6; age²: 8.6), with a similar pattern to that observed in the early time window: a decrease in grand average amplitude with age, following a slightly concave trajectory. Although the contrast $\times$ age² interaction reached statistical significance (BSR = -3.2), the corresponding effect (i.e., reduced condition differences at the youngest and oldest age) was modest and not clearly visible in the data (Figure 5B-C). Additional data will be needed to assess the robustness and replicability of this pattern. The same PLS-c in the sleeping subsample yielded no significant latent component, suggesting that sleep may reduce or even abolish the late response to word novelty.

A PLS-c analysis of the difference part-words *minus* words using group as contrast revealed one significant latent component ($p=.002$; $r=.65$; 29.3% explained covariance, Figure 6), with a similar topography and timing as the latent component identified on the activation using words vs part-word as contrast (Supplementary materials, Figures S7-8 show grand average topographies per group and condition). This latent component included a significant group effect, with stronger frontal negativity and posterior positivity in LL participants compared to HL (BSR: 15.7). Additionally, significant negative age (BSR: -5.8), age² (BSR: -3.7), and age²*group interaction (BSR: -12.1) effects were observed. Furthermore, the latent component included a positive effect of both verbal outcome (BSR: 16.1) and verbal outcome*group interaction (BSR: 3.0). This indicates that the magnitude of the late response to novelty predicted verbal outcomes. Given that no late response was detected in sleeping participants, we re-ran the PLS-c analysis using group as a contrast in the awake subsample ($n=58$). This yielded one significant latent component ($p=.006$, $r=.68$, 30.5% explained covariance), with behavioral and electrode contributions highly overlapping with those in Figure 6.

As we did for neural entrainment to syllables, we conducted a new analysis on late ERP to word novelty, excluding EEG data from the final visit. This PLS-c yielded one significant latent component ($p=.002$, $r=.74$, 33.5% explained covariance, Figure 6-figure supplement 1) with globally similar EEG parameter contributions and age trajectory modelling. Verbal outcome still significantly contributed to the latent component (BSR: 20.1), with a negative verbal outcome*group interaction (BSR: -7.2). These results suggest that, after ruling out cross-sectional correlations at 18–21 months, the late ERP to word novelty predominantly predicts verbal outcomes in high-likelihood infants for autism.

Given that PLS-c analyses in the late window revealed both condition effects (words vs. part-words) and group differences on overlapping latent components, we asked whether the observed effect was present in both groups or driven primarily by LL participants. To test this, we ran separate supplementary PLS-c analyses within each group using the part-word vs. word contrast (see Figure 6-figure supplement 2-3). Both analyses revealed a significant latent component ($p<.001$), with similar spatial topographies and temporal profiles. Crucially, the condition contrast

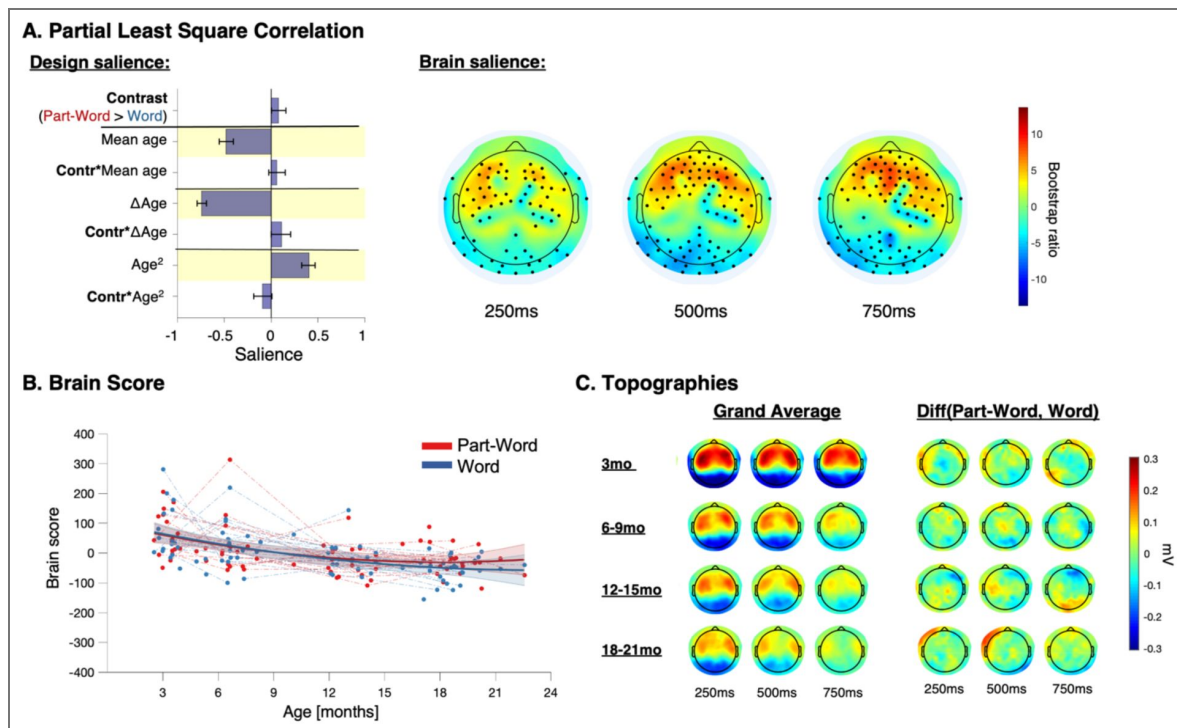


Figure 4. Early evoked response potential (ERP) to part-words compared to words across all participants.

A. Design and brain saliences derived from the significant latent component. Brain topographies of bootstrap ratios (BSR) are displayed at 250ms intervals. Black dots indicate BSR > 2.3. **B.** Participants' brain scores for part-word and word conditions, as a function of age. Brain scores are participants' raw voltage data projected onto electrode saliences. Brain scores illustrate how individual EEG data fit the saliences derived from the latent component. Linear fitting is used for illustrative purposes only. **C.** Voltage grand averaged (left) and differential (right) responses to part-word and word conditions at each age bin: 3 months (n=18), 6-9 months (n=20), 12-15 months (n=20), 18-21 months (n=25).

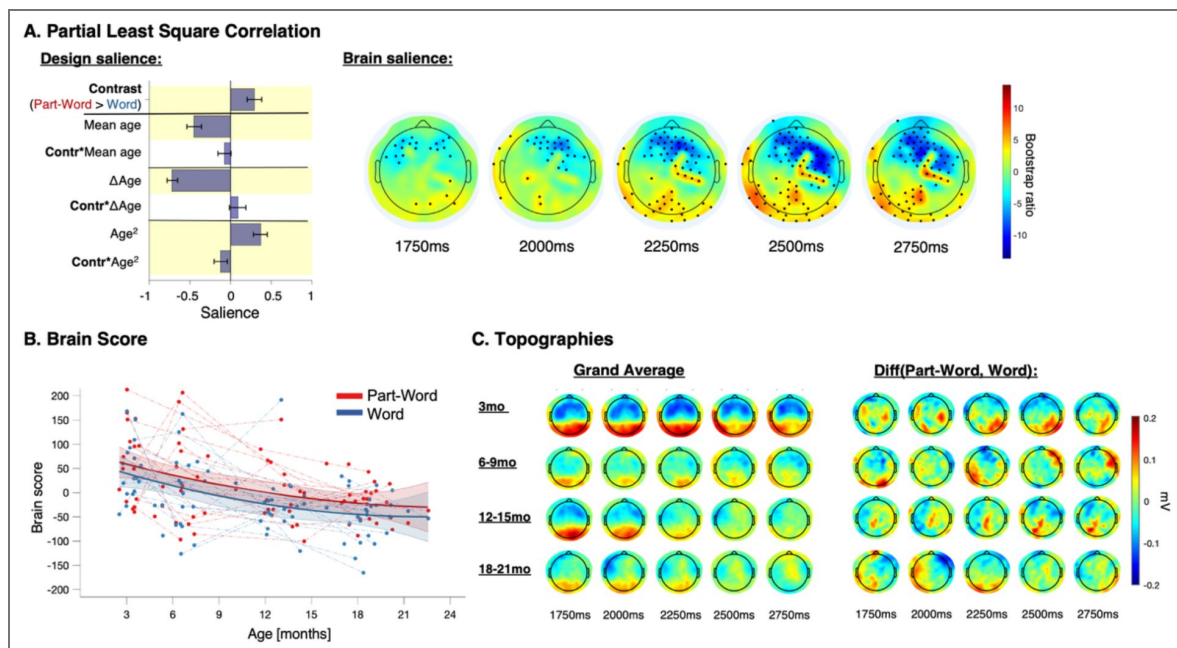


Figure 5. Late evoked response potential (ERP) to part-words compared to words across all participants.

A. Design and brain saliences derived from the significant latent component. Brain topographies of bootstrap ratios (BSR) are displayed at 250ms intervals. Black dots indicate BSR > 2.3. **B.** Participants' brain scores for part-word and word conditions, as a function of age. For details, see figure 4. **C.** Voltage averaged (left) and differential (right) responses to part-word and word conditions at each age bin: 3 months (n=18), 6-9 months (n=20), 12-15 months (n=20), 18-21 months (n=25).

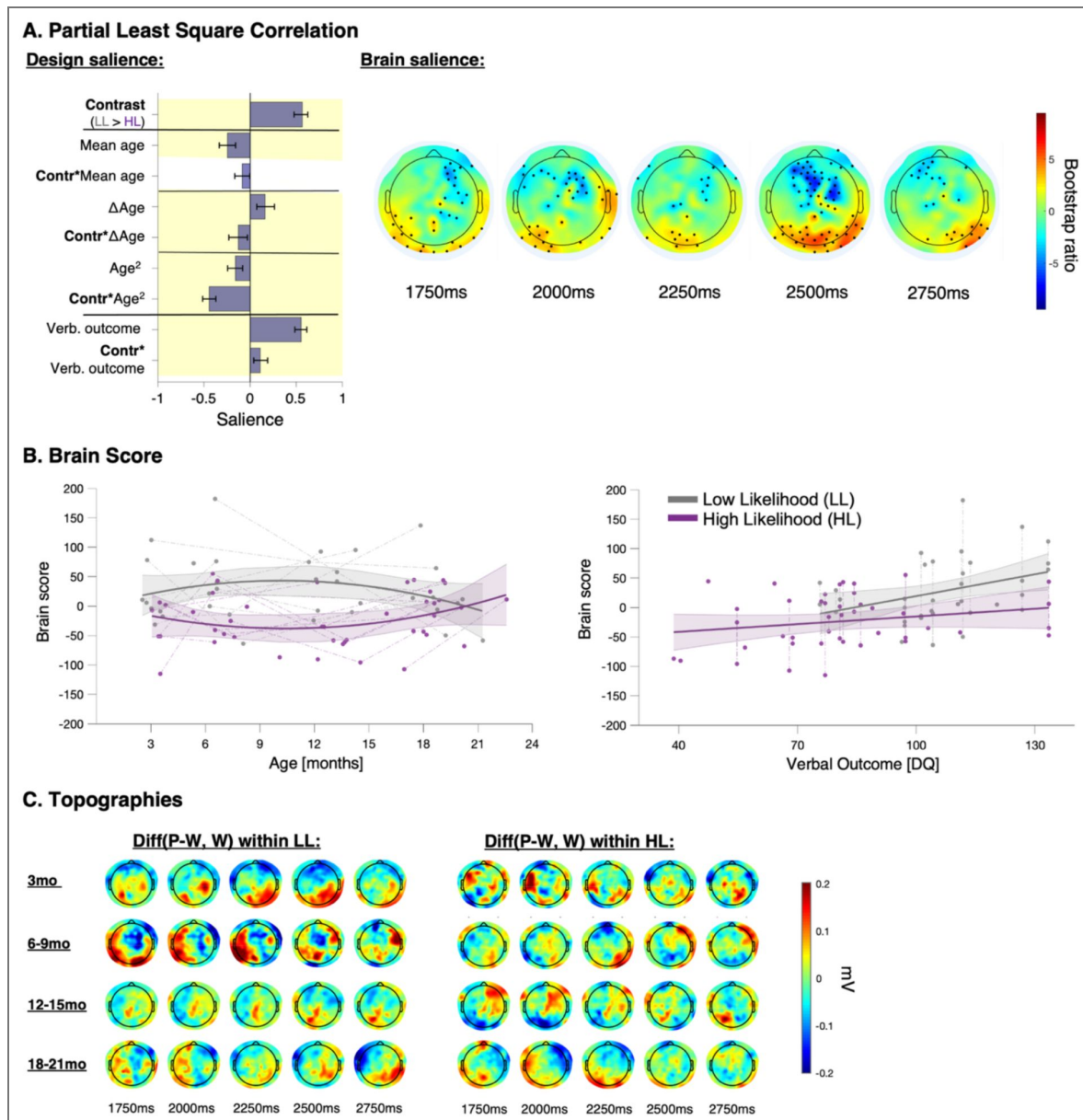


Figure 6. Late evoked response potential (ERP) to novelty in infants at high and low likelihood for autism.

A. Design and brain saliences derived from the significant latent component. Brain topographies of bootstrap ratios (BSR) are displayed at 250ms intervals. Black dots indicate BSR > 2.3. **B.** Participants' brain scores for part-word and word conditions, as a function of age (left) and of verbal outcome (right). Detailed information about brain scores is provided in Figure 4. Linear fitting is used for illustrative purposes only. **C.** Voltage differential responses to part-word and word conditions at each age bin and within each group: 3 months (11 LL and 7 HL), 6-9 months (9 LL and 11 HL), 12-15 months (10 LL and 10 HL), 18-21 months (9 LL and 16 HL). DQ: developmental quotient. HL: high likelihood for autism; LL: low likelihood for autism.

contributed significantly only in the LL group (BSR: 13.0), but not in the HL group (BSR<2.3), suggesting that the group difference observed in Figure 6 reflects the absence of a detectable novelty response in HL participants with the current analysis.

Overall, we found no difference between conditions during the early time window, which was aligned with the temporal progression of syllables within the triplet ([0–1000] ms; Figure 4). However, in a later time window ([1500–3000] ms) following triplet onset, the ERP exhibited a differential response to part-words and words (Figure 5). This late effect was positively associated with verbal outcomes at 20 months and, notably, was not detected in the HL group.

Discussion

We exposed 44 infants at low (LL) and high (HL) likelihood for Autism Spectrum Disorder to an artificial speech stream composed of syllabic triplets (words) with random speaker changes at every syllable to assess their ability to track both syllables and word-level structure under conditions of acoustic variability. We focused on three electrophysiological measures: 1) neural entrainment to syllables; 2) neural entrainment to words; and 3) ERP responses to isolated words and part-words, analyzed in early [0–1000ms] and late [1500–3000ms] windows. These analyses allowed us to trace the sequential stages of speech processing, from syllabic perception to online word segmentation and later word recognition, and to identify the levels at which HL infants may encounter difficulties. While the core mechanism of statistical learning appeared intact in HL infants, as evidenced by word-level entrainment in both groups, several group differences emerged in the dynamics of syllable tracking and in the late ERP responses. Notably, some of these effects were significantly related to verbal outcomes at 18–21 months.

Syllable neural entrainment

HL infants exhibited significantly weaker syllable-level neural entrainment than their LL peers (Figure 2), despite following a similar convex developmental trajectory across the first two years of life. This persistent lag in rhythmic tracking was not only stable over time but also strongly predictive of verbal outcome at 20 months, highlighting its potential as an early neurophysiological marker of language development. This result refines previous observations regarding reduced theta power in HL infants and autistic children exposed to natural speech (28,29). Unlike these studies, which estimated speech-brain coherence across the theta band, our analysis specifically targeted the particular frequency at which the auditory sequence was presented (4 Hz) and isolated phase-locked activity by subtracting adjacent frequency components. This more selective approach points to a precise impairment in synchronization with syllabic units, rather than a general reduction in low-frequency neural activity, indicating a targeted disruption of syllabic tracking.

Interestingly, the divergence between HL and LL infants emerged specifically during the structured stream and coincided with the time window in which infants began to segment the continuous speech flow into word-like units (Supplementary Material Figure S4–5). This temporal overlap suggests that successful segmentation may enhance syllable tracking via top-down predictions of the next syllable, improving alignment to syllable onsets in LL infants as well as in HL with better verbal outcome.

While this top-down interpretation highlights the role of word segmentation in enhancing syllable tracking, it does not exclude additional contributions from lower-level sensory mechanisms. According to the oscillatory framework of speech perception, syllable tracking is thought to be supported by endogenous neural oscillations in the theta band, which align with the natural rhythm of speech (54). Disruptions in this mechanism have been proposed in autism (28,29). However, recent evidence suggests that this model may not fully apply to early infancy. In typical 3-month-old infants, we observed robust neural entrainment to amplitude-modulated sounds across a wide frequency range (2–45 Hz), with more accurate tracking below 12 Hz — but no selective enhancement in the theta range (Kabdebon et al., submitted), challenging the idea of particular sensitivity to entrainment in this frequency range, at least in infants.

Beyond oscillatory alignment, another possibility is that weaker syllable tracking in HL infants stems from temporally and/or frequency-imprecise responses along the auditory pathway. Impaired temporal or spectral fidelity, particularly at the brainstem level, has been reported in autism and other language-related developmental disorders (37,55). In this context, decreased cortical power at specific frequencies observed in dyslexic adults (56) and autism (57) might reflect the long-term consequences of lower-level auditory processing difficulties rather than their root cause. Supporting this view, recent findings show that atypical brainstem responses can be detected even in first-degree relatives of autistic individuals, and that lower brainstem response consistency is associated with poorer pragmatic language skills in these individuals (58). These results highlight the cascading effect of low-level auditory processing anomalies on higher-level verbal and communication abilities. Such a cascade could explain the observed relationship between syllabic entrainment and verbal outcomes at 18–21 months. Future studies should directly assess whether degraded brainstem responses underlie reduced entrainment in HL infants using higher EEG sampling rates.

In sum, HL infants show a specific and developmentally stable deficit in syllable tracking, which may reflect reduced top-down support from word segmentation, impaired oscillatory alignment, and/or imprecise encoding in the brainstem. Crucially, this neural signature was predictive of later verbal outcomes, underscoring its potential relevance as both a mechanistic insight and an early clinical marker. We now turn to word entrainment results, which offer a complementary window into how infants begin to extract and consolidate word units from speech.

Word segmentation: neural entrainment

Neural entrainment at the word rate (1.3 Hz) provides a robust index of speech segmentation: it can only emerge once infants begin grouping syllables into coherent word-like units. Despite the acoustic variability introduced by speaker changes, both HL and LL infants showed significant word-level entrainment after 90 seconds of exposure to the structured stream (Supplementary Material Figure S3 [3](#)), replicating previous findings with single speakers (15,16). Although HL infants exhibited reduced syllable tracking, no overall group difference was observed for word entrainment (Figure 3 [3](#)). However, the low frequency of the word rhythm, coupled with high power in the same range in infant EEG, may reduce the signal-to-noise ratio, limiting the sensitivity of this comparison.

Our longitudinal design, however, revealed a striking divergence in developmental trajectories. LL infants displayed a U-shaped curve in word entrainment, with strong responses at 3 and 21 months and a dip around 12 months (Figure 3D [3](#)). In contrast, HL infants showed a flatter trajectory, with consistent performance across ages. Given the scarcity of longitudinal studies on statistical learning in infancy (42), these data provide novel insight into the developmental dynamics of word segmentation. Importantly, previous studies have shown that statistical learning is present from birth (14,15,44), and that word learning trajectories in 6-month-olds resemble those observed in adults (16), suggesting that this 12-month dip does not reflect a lack of statistical learning ability, but rather a transient disruption of an otherwise early and robust mechanism.

The dip in LL infants around 12 months coincided with an increase in syllable entrainment (Figure 2B [3](#)), making it unlikely to result from reduced data quality or measurement artefacts. We propose, as a post-hoc hypothesis, that the random changes in speaker identity between syllables disrupted word segmentation. In typically developing infants, voice processing performance improves toward the end of the first year (59,60), alongside the accumulation of implicit knowledge about how speech is typically structured — notably, that words are usually produced by a single speaker and that they do not overlap prosodic boundaries. These emerging priors may constrain the segmentation process by restricting it to plausible word units, as has been shown for prosodic boundaries (61). In our paradigm, where speaker identity changed randomly between each syllable, this conflict may have temporarily disrupted this process, leading to a reduction in power at the word frequency.

Simultaneously, growing attention to voice-related features may have improved alignment to syllables, consistent with the convex concave shape of syllable entrainment. The 12-month dip in word tracking may also reflect the attentional cost of reorienting to each new voice (62,63), during a period when infants are intensely engaged in learning social cues. This interference appears to be transient: word segmentation recovers at later ages, potentially as infants become more efficient at managing speaker variability and less disrupted/interested by voice changes. This trajectory aligns with adult findings showing that word segmentation in continuous speech is context-sensitive, shaped not only by input statistics but also by the relative weight of different priors (12–50).

HL infants, on the other hand, did not show this transient disruption. In this group, word entrainment remained stable over time. To account for this unexpected finding, we followed up on the post-hoc hypothesis proposed above: a reduced sensitivity to social and vocal cues observed in HL infants may have spared segmentation abilities by limiting the interference introduced by speaker variability (64). If this post-hoc hypothesis holds true, LL and HL infants would differ not in their intrinsic ability to learn statistical regularities *per se*, but rather in how they integrate or suppress competing cues (such as speaker changes) during the segmentation process. It is important to note, however, that the present study was not designed to isolate and evaluate the specific impact of speaker changes on word segmentation. Consequently, this interpretation remains speculative, and additional research is required to further address this question.

While HL infants may paradoxically benefit from a reduced sensitivity to conflicting social cues in our specific paradigm, this same attenuation could become a disadvantage in more naturalistic settings, where voice identity supports word recognition, such as attributing speech to the correct speaker during a multi-party conversation. Nonetheless, this interpretation remains speculative in the absence of a control condition with a consistent speaker. Future studies manipulating speaker variability will be essential to determine its causal role in shaping the divergent developmental trajectories observed between groups.

Word recognition: ERP

In the test phase, we observed a robust late ERP response to novel pseudo-words compared to familiar triplets, emerging around 750 ms after word offset (Figure 5). This response reveals participants' ability to recognize the words they have been exposed to during the stream. It belongs to the family of late slow-wave components observed in infants, which have been associated with higher-order novelty detection and orientation mechanisms (53). Crucially, the amplitude of this late response correlated positively with verbal outcomes at 20 months, suggesting that the ability to track and later recognize newly learned words may serve as an early predictor of language development. As with syllable entrainment, the late ERP to novel words primarily predicted verbal outcomes in high-likelihood (HL) infants.

Despite their transient drop in word tracking at 12 months, LL infants still exhibited a difference between words and part-words at test, indicating that they had successfully memorized the transitional probabilities between syllables. In other words, voice-related interference during the learning phase seems not to prevent participants from recognizing a familiar word form in test. Statistical learning can thus remain intact even when evidence for online segmentation is weak or absent — a dissociation previously reported in neonates and adults (43).

In contrast, HL infants showed no clear ERP difference between novel and familiar triplets (Figure 6-figure supplement 3). Given that both groups showed similar word neural entrainment during learning, this absence is unlikely to reflect a failure in statistical learning itself. Instead, it may point either to a deficit in novelty orientation — a phenomenon previously described in autism and HL infants (65,66), or to a failure to explicitly recognize the novel part-word. A recent study in adults showed that implicit and explicit traces of statistical learning rely on separate consolidation mechanisms, with implicit learning being more robust (67). Interestingly, this dissociation between spared implicit versus impaired explicit statistical learning in autism has been previously proposed and discussed in the literature (68,69). According to these studies,

autistic impairments in top-down attentional processes, such as social orienting — which are critical for bootstrapping language acquisition (70) — may result in a heightened dependence on bottom-up mechanisms, including implicit statistical learning.

Finally, we found no early ERP distinction between part-words and words in the [0–1000 ms] window following word onset (Figure 4B). This contrasts with findings in neonates, who rely on the first-syllable to recognize recently learned words (15). This result suggests that by 3 months of age, infants may shift from relying on recognition of initial syllables to encoding entire word forms, a more mature form of lexical storage and retrieval. Alternatively, the topology of early responses might change during development, making the identification of these patterns difficult, given the relatively small number of participants tested at each age.

Implications for early detection and intervention

Our results highlight several neurophysiological markers that may prove useful for early identification of infants at heightened likelihood for language difficulties or neurodevelopmental disorders. Importantly, these measures rely on passive EEG paradigms, making them accessible, non-invasive, and feasible even in very young or at-risk populations.

Syllable neural entrainment, which was consistently weaker in HL infants and predictive of later verbal outcomes, may serve as an early indicator of atypical speech tracking. This deficit may disrupt the ability to process and integrate phonetic and phonotactic cues critical for internalizing the rules of the native language. Moreover, interventions known to enhance auditory encoding — such as music-based training — have shown benefits for brainstem precision and language outcomes (71,72), and may prove particularly valuable in at-risk populations.

Likewise, the absence of a late ERP orientation response in HL participants may represent an early neural signature of altered attention to novelty that can be used both as a non-invasive predictor of language development and as a potential target for early intervention. This potential biomarker might nevertheless be modulated by participants' sleep status, warranting careful consideration of vigilance state in future studies.

Limitations

Several limitations of this study should be acknowledged. First, given the multidimensional nature of the EEG data and the longitudinal design, we relied on Partial Least Squares Correlation (PLS-c) analyses to identify latent components linking neural responses with experimental and developmental variables. While this multivariate approach was necessary to reduce data dimensionality and handle collinearity, it captures only shared variance across participants. As a result, PLS-c may miss more localized or subtle effects that do not align with the dominant latent structures. Thus, although our longitudinal design offers valuable insights into developmental trajectories, the uneven age distribution and limited number of repeated measures per infant constrain the interpretation of individual growth curves. More densely sampled longitudinal data would be needed to precisely model intra-individual changes. Second, although all participants were primarily exposed to French, we did not quantify additional language exposure, precluding any analysis of its potential moderator effects on statistical learning in our groups and age-trajectories. However, prior work has reported no effect of bilingualism on auditory triplet segmentation in children (73). Finally, while we interpret reduced syllable entrainment and diminished novelty responses in HL infants as potential early markers of later verbal outcomes, the predictive validity and autistic specificity of these neural indices should be confirmed in larger cohorts and across more diverse developmental profiles.

Conclusion

Our study underscores the value of longitudinal neuroimaging—both online and offline—for unpacking the hierarchical processes underlying infant speech processing in both HL and LL populations. Even in typical development, these mechanisms remain poorly characterized. By

mapping their developmental trajectories across the first two years of life, we revealed how early speech processing abilities are not static, but dynamically shaped by concurrent maturational changes.

Importantly, our results illustrate how subtle low-level differences—such as reduced syllable tracking or diminished novelty responses—can cascade into broader developmental outcomes. This supports a neuroconstructivist perspective (74), where early neural variability may help explain later divergence in language and communication, both in typical and atypical pathways.

Materials and Methods

Participants

Forty-four infants from the ongoing Geneva Autism Cohort were included during longitudinal visits, contributing to 83 EEG recordings (4,75). The open longitudinal design comprised four longitudinal visits: (1) at 3 months, (2) between 6 and 9 months, (3) between 12 and 15 months and (4) between 18 and 21 months. Verbal outcome was collected at the last visit (18-21 months). Longitudinal neuroimaging designs reduce within-subject variability by distributing the information across time (76,77). All participants were raised in primarily French-speaking environments, with French as the dominant language at home and daycare. The standardized developmental assessment Mullen Scale of Early Learnings was administered at the last visit to estimate participants' verbal developmental quotient (DQ), our outcome measure (78). Participants' developmental ages at the 18-21 months visit were divided by their chronological age to get developmental quotients (DQ) centered on 100. DQs avoid floor-effects by not truncating the very low-performing participants' scores (79,80). The current study was approved by the University of Geneva's Ethics Committee, and parents provided written informed consent. Sample characteristics are provided in Table 1 (81).

The 19 LL infants were healthy full-terms (>37 weeks of pregnancy) without any reported autism in their 1st degree relatives. From the 25 HL infants, 18 had an older autistic sibling (n=18), which is known to be associated with a 18% prevalence of autism (81). The 7 other HL infants presented with early parental concerns for autism, based on parental report prior to enrollment. Their Autism Parent Screen for Infants (APSI) total score at their 18-21 months visit was 15.6±6.4, [8-22] range – a score greater than 8 reflecting a 63% positive predictive value for autism in HL populations (82). One of the HL participants with early parental concerns had a PACS1 mutation, a condition associated with a 37% autism prevalence (83). Moreover, 4 HL infants were born preterm (range: [31-36] gestational weeks), a condition associated with a 7% autism prevalence (84). Corrected age was used by subtracting the number of prematurity weeks from the chronological age (78,85).

There was no statistically significant difference (chi-square, $p>.05$) between HL and LL in either biological sex, visits' repartition, or age (Table 1 (81)). Verbal outcome was missing for 4 participants (3 LL and 1 HL, 4 recordings) because of drop-out before the 18-21 months outcome visit. Those participants were excluded from analyses using verbal outcome as a parameter.

Of the initial sample (103 EEG recordings), 20 acquisitions from 16 participants (5 LL and 11 HL) were not included (19.4% data loss), due to too many crying and/or motion artifacts after visual inspection of the data (3 LL and 13 HL recordings), or to examiner's acquisition error (2 LL and 2 HL recordings).

Stimuli

Stimuli and procedure were based on Fló et al's paradigm (15,44). Twelve syllables were synthesized with MBROLA (86) using French phonemes and phonotactic rules. Each syllable lasted 250 ms, had a flat intonation, and was produced by 3 male and 3 female MBROLA speakers. We built two different stream conditions: a structured stream (STR) and a random control stream (RND) by concatenating the syllables with speakers varying randomly from one syllable to the next. This created a rhythmic syllabic presentation at 4Hz in both streams. In the RND stream, syllables were presented randomly, maintaining a flat transition probability of 1/11 between

MEASURE [mean (SD)]	Low-Likelihood	High-Likelihood	P-Val
Number of participants (n=44)	19	25	
Number of EEG recordings (n=83)	39	44	
3 months EEG age (n=18)	3.3±0.6 (n=11)	3.5±0.3 (n=7)	.309
6-9 months EEG age (n=20)	6.6±0.8 (n=9)	7.1±1.3 (n=11)	.320
12-15 months EEG age (n=20)	13.0±1.0 (n=10)	13.5±1.4 (n=10)	.371
18-21 months EEG age (n=25)	19.0±1.1 (n=9)	18.7±1.6 (n=16)	.438
Female biological sex	4 (21.1%)	12 (48.0%)	.066 (χ^2)
Age at verbal outcome	19.6±2.4 (n=16)	19.7±2.0 (n=24)	.886
Verbal outcome [DQ]	104.3±16.2	77.9±22.2	.001

Table 1. Sample characteristics.

Statistical comparison between LL and HL samples. For categorical variable, chi square (χ^2) was applied. For continuous variables, we used two-tailed independent T tests. P values < .05 are highlighted in bold.

syllables. In STR, syllables were organized into 4 tri-syllabic words presented in random order but without repetition, resulting in transition probabilities equal to 1 within words and a drop in transition probability to 1/3 at word boundaries. Words were paced at 1.33 Hz ($1/(3 \times 0.25s)$). We built two RND streams lasting 90 s each, one long STR stream lasting 180 s, and 6 short STR streams lasting 30 s each. Note that the voice dimension was totally orthogonal to the present design, each syllable being randomly produced by one of the six speakers. Thus, 1) the voice was not constant within a word and 2) there was acoustic variability between words (i.e. different voices produced the same syllables). This implies a voice normalization process to recognize the same syllables and words across different occurrences. More details about the stimuli are provided in the Supplementary Material, Stimuli section.

To explore infants' word recognition after familiarization with the STR stream, we constructed isolated triplets corresponding to *words* and *part-words*. *Words* exactly matched the triplets used to build the STR streams ($A_i B_i C_i$ structure, letters being syllables from the i th learned word). *Part-words* included the 2 last syllables of a word i , with a random first syllable of another word k ($B_i C_i A_k$ structure). In a *part-word*, the transition probability between the two first syllables was 1, but these syllables were not in the correct position in the word. Furthermore, the transition probability between the second and the last syllable was 0.33 instead of 1. While a late ERP difference would signal that infants were sensitive to any of this information, an early difference would signal participants' recall of the ordinal position of the syllables within the learned words (Flò et al, 2022).

Data Collection

We collected high-density electroencephalograms (EEG) with a 128-electrode net (Electrical Geodesics, Inc) referenced to the vertex. The sampling frequency was 250 Hz. Participants were sitting on their parent's lap, and a silent cartoon was presented to keep them quiet and still. Stimuli were played on a Bose® Companion 2 Series III at a 50cm distance with an intensity of 75dB. The cartoon was not time-locked to the auditory stimuli and varied across participants. The same examiner (MG) collected all data. When participants were restless and/or not interested in the cartoon, the examiner engaged them in quiet activities (e.g., staring at bubbles/toys). Some EEGs were recorded while participants were asleep (61.1% at 3 months, 40% at 6-9 months, 20% at 12-15 months and 4% at 18-21 months). There was no significant difference ($p=.797$) in sleeping status between HL and LL (linear mixed-effect model with repeated measures). Neural entrainment and statistical learning have been shown to be present in sleep in adults and neonates (15,16). The experimental procedure is detailed in Figure 1A [\[15\]](#).

Data pre-processing

Data were first resampled to 300 Hz to get an integer sample number in each three-syllabic item, then band-pass filtered (0.2-40Hz). APICE preprocessing pipeline (https://github.com/neurokidslab/eeg_preprocessing [\[17\]](#)) was applied using EEGLAB toolbox 2020.0 on Matlab® R2018b (87,88). In brief, bad segments of data were identified using algorithms detecting low correlation with other channels (usually due to non-functional channels) and outlier values for the signal amplitude or changes in the signal amplitude (typically due to motion artifacts), with a threshold of 2 (outliers are values bigger than two interquartile ranges away from the third quartile). A sample was considered to contain motion artifacts if more than 30% of the working electrodes were rejected. An electrode was considered bad if more than 30% of the free-of-artifacts samples were rejected. Short rejected periods (less than 100 ms) were corrected using target PCA. In segments containing less than 30% of the electrodes marked as bad artifact data were spatially interpolated using spherical spline interpolation. A careful data visual inspection was also carried out with manual removal of remaining bad electrodes and motion artifacts (done by MG). Independent Component Analysis (ICA) and the iMARA algorithm for component classification (89) were used to remove physiological noise. Finally, bad electrodes were interpolated using spherical splines.

A) Neural entrainment analyses

Given the construction of the streams with regular stimuli, we expected a neural entrainment at the stimuli specific frequencies (syllabic rate for the STR and RND streams and word rate when words were discovered in STR). This specific entrainment should not be observed during rest. We thus measured neural entrainment at 4 Hz (syllabic rate) and 1.3 Hz (word rate) in each period and each recording. We did so by considering non-overlapping epochs of 7.5 seconds respecting the chronological order of presentation (90). Epochs contaminated by artifacts were excluded from the analysis. On average, we kept 86% of data during RS (range of included epochs: [11-24]), 88% during RND ([12-24]), and 91% during STR ([32-48]). The number of included epochs did not differ between LL and HL groups, nor vary with age, suggesting an even data quality across groups and age (see Data quality section in Supplementary Material, S1 (3)).

Each epoch was average-referenced and normalized by dividing by the standard deviation computed across all electrodes and samples of each epoch. Denoising source separation (DSS) was applied to remove stimulus-unrelated activity using spatial filtering (91). Briefly, a PCA was first applied to the epoched data. Afterward, a DSS filter was applied on the 30 first components, and its 6 first components were kept. Eventually, a fast Fourier transform (FFT) was applied to the denoised epochs to estimate their phase-locking value (PLV). PLV estimates how much EEG activity is synchronized in phase with a specific frequency (1):

$$PLV(f) = \frac{1}{N} \left| \sum_{k=1}^N e^{i\varphi(f,k)} \right| \quad (1)$$

With N being the number of trials from one stream and $\varphi(f, k)$ the phase at frequency f and trial k . PLV ranges from 0 (desynchronized activity) to 1 (phased locked activity). PLV was computed in 31 frequency bins (0.933 - 4.933 Hz, with 0.133 increment). In each stream, the PLV of each frequency bin was Z-scored over the 12 adjacent frequency bins (six on each side)(2):

$$\text{Z-scored PLV}(f_a) = (\text{PLV}(f_a) - \text{mean}(\text{PLV}(f_x))) / \text{std}(\text{PLV}(f_x)) \quad (2)$$

Where f_a is the targeted frequency bin, and $\text{PLV}(f_x)$ is the PLV over the 12 adjacent frequency bins. From now on, PLV refers to Z-scored PLVs. Additionally, in each electrode, PLV values from immediate adjacent frequencies (in which no entrainment is expected) were subtracted to get a cleaner signal (e.g., 1.2-1.47Hz for 1.3Hz word rate and 3.87-4.13Hz for 4Hz syllable rate).

Longitudinal Partial Least Square Correlations

Partial Least Square Correlation (PLS-c) is a multivariate statistical approach that has been successfully implemented for EEG data (50,92,93). PLS-c can also be applied to longitudinal neuroimaging datasets, including a variable number of timepoints per participant, as ours (51,94). In a fully data-driven approach, one single longitudinal PLS-C can detect age-trajectories of electrophysiological measures (here, PLVs), indicate the electrode clusters in which those trajectories take place, and find their association with behavioral parameters (here, HL/LL group and verbal outcome). We applied longitudinal PLS-c based on the pipeline described in Delavari et al. 2021 using the myPLS toolbox (95,96) : <https://github.com/danizoeller/myPLS> (3). Briefly (Figure 1B (3)): using each participant i 's n^{th} visit, we built a behavior design matrix (Y^T) with 9 variables: 1) Contrast (a binary contrast, in the figure example corresponding to LL or HL group), 2-3) two orthogonalized variables to grasp the cross-sectional and longitudinal effects of age (Mean age = averaged age across all n visits of the participant i , and Delta age = difference between participant i 's age at visit n and the participant i 's Mean age), 4) another age variable to grasp convex/concave trajectories ($\text{Age}^2 = \text{Delta age} * [\text{Mean age averaged across all participants}]$), 5-7) three interaction variables: Mean-age*Contrast; Delta-age*Contrast; $\text{Age}^2 * \text{Contrast}$, 8-9) Verbal outcome (DQ collected at the 18-21mo visit) and one interaction variable: Verbal outcome*Contrast. We limited the number of behavioral variables to nine to mitigate noise sensitivity and overfitting risks associated with exceeding the 10% sample size threshold (97). This limitation precluded the implementation of a single PLS-c model incorporating group, condition (STR vs. RND), age, and their interactions. Behavior design variables were z-scored across all participants. We also built a

brain data matrix (X) for each participant i 's visit n including the 128 electrodes PLVs (1.3 Hz in STR when testing entrainment to word rate, and 4 Hz in both STR and RND when exploring entrainment to syllable rate). We then computed cross-variance matrices (R) as $Y^T X$. R underwent singular value decomposition ($R = USV^T$) to derive 9 singular values called latent components. Permutation testing (1000 permutations shuffling behavioral data across participants) estimated whether each latent component statistically significantly explained the correlation. Bonferroni correction was applied to account for multiple comparisons across the 9 tested latent components in the PLS-c, yielding an adjusted alpha of .006 (51,98). In each significant latent component, bootstrapping (500 random samples and replacement) was applied to evaluate the saliency of each behavior and brain (electrodes' PLV) variable, reflecting the stability of its contribution to the latent component. Saliencies are summarized as bootstrap ratio (BSR): mean of bootstrap results divided by standard error. BSR are analogous to Z-scores and can be used to assess the stability of the saliency. We considered $BSR > 2.3$ as stable, corresponding to a 99.0% bootstrap confidence interval not crossing zero – roughly equivalent to a two-tailed $p < .001$ (50,51). To respect the intra-participant longitudinal dependencies, the bootstrap samples were randomly selected across participants and not EEG recordings. To help readers interpret the complex output of the PLS-c, we created plots of the individual raw EEG data extracted from the salient electrodes as a function of key behavioral variables (e.g., age, age-squared, and verbal outcome). We overlaid a polynomial curve (fitted using a least-squares mixed-effects model) on these raw data plots for visualization purposes only. This curve is strictly intended to assist with interpretation and is not meant as an additional statistical analysis, since the statistical relationship between EEG and behavioral variables was already established by the PLS-c.

It is important to note that PLS-c identifies EEG spatial patterns shared across ages and groups, allowing us to measure group and age effects within these shared spatial maps. Consequently, this method may overlook subtle spatial differences that could exist between specific groups and age categories.

B) Analyses of ERPs to test items

The preprocessed data were low-pass filtered (20 Hz) and epoched between $[-1.75, 3.25]$ s from the triplets' onset. Epochs containing artifacts were excluded. On average, we kept 83% of data for Words (range of included epochs: [23-48]), and 83% for Part-Words ([19-48]). Included epochs tended to decrease with age, but did not differ between LL and HL groups, and there was no age*group interaction, suggesting a similar data quality across groups (Supplementary Material, Figure S2). Each participant's data was reference-averaged and normalized by dividing by the standard deviation computed across all electrodes and samples of each epoch. Trials were averaged by condition (Words and Part-words).

ERP longitudinal Partial Least Square Correlations

We applied longitudinal PLS-C on ERP data to investigate the trajectories of word recognition between LL and HL, using the same method as above with the following differences: brain matrices (X) included voltage measures from each electrode at each sample (3.33 ms time frame) instead of PLVs at each electrode. Consequently, X were time*space matrices (50). Moreover, to summarize the ERPs relationship with age and squared age, we computed brain scores. This metric reflects the projection of participants' raw voltage values in the electrode saliencies obtained from the PLS. Brain scores thus provide summary values reflecting how well each EEG acquisition fits the brain saliencies obtained by a given latent component. Fitting a mixed-effect model on brain scores allowed an illustration of the ERP trajectory with age. The same limitations outlined earlier apply to PLS-C when using ERP data — namely, it identifies spatial patterns shared between groups and age bins, which may lead to missing subtle, group- or age-specific differences.

Figure supplements

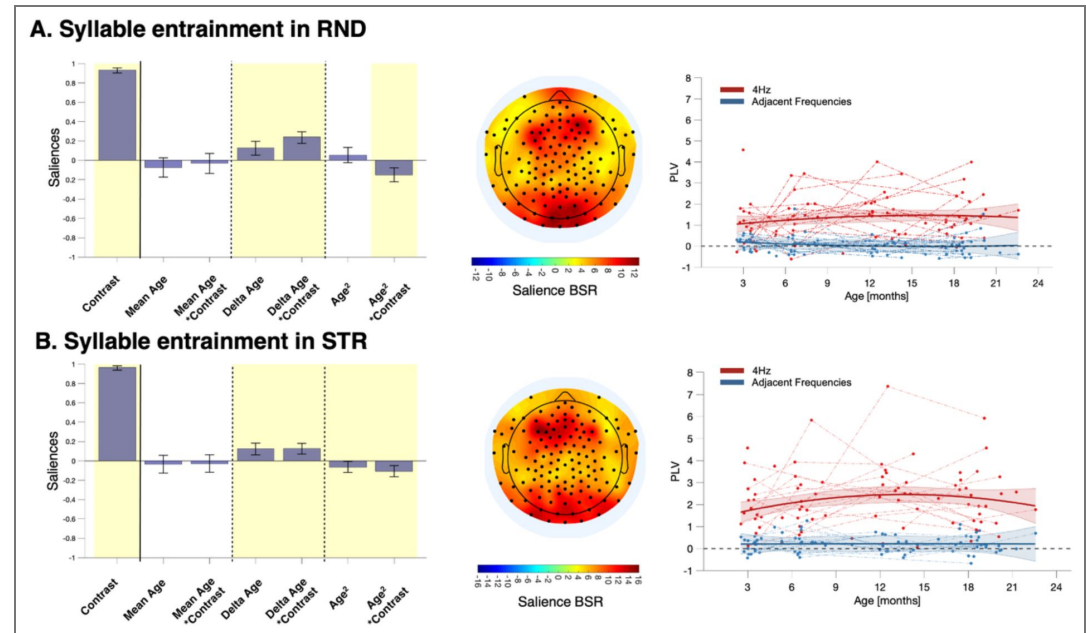


Figure 2-figure supplement 1. Follow-up analyses restricted to each stream for syllable entrainment. A. For RND, the latent component was significant ($p < .001$; $r = .67$; 87.3% explained covariance) with the following BSRs: contrast : 70.3*; mean age: -1.6; contrast*mean age: -0.7; delta age: 3.6*; contrast*delta age: 8.0*; age²: 1.2; contrast*age²: -4.1*. **B.** For STR, the latent component was significant ($p < .001$; $r = .73$; 91.1% explained covariance) with the following BSRs: contrast : 91.4*; mean age: -.9; contrast*mean age: 0.7; delta age: 4.2* contrast*delta age: 4.7*; age²: -2.4; contrast*age²: -3.8*. Yellow shading on left panels and black dots on middle panels indicate BSR > 2.3 – BSR > 2.3 is considered significant.

Figure 2-figure supplement 2. Neural entrainment to syllable rate in sleeping participants (n=25 recordings).

A. Design salience (left) and brain salience (right topography) derived from the significant latent component for neural entrainment to the syllable rate using frequency (4Hz versus adjacent frequencies) as contrast. Significance (i.e., salience) was established through bootstrapping. **B.** Individual raw Phase Locking Values (PLVs) extracted from the salient electrodes identified by the latent component (black dots in the topography on 2A) are displayed. A fitted curve is included for visualization purposes only, produced by a mixed-effects model with a 95% confidence interval.

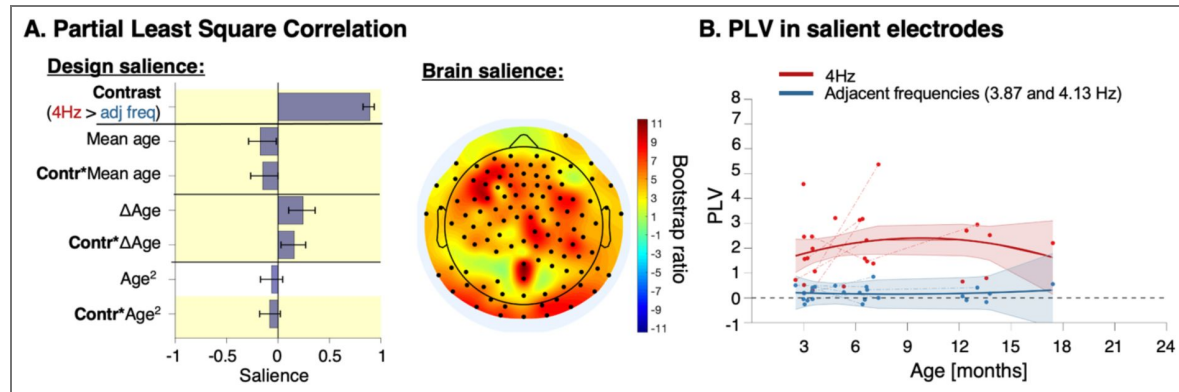


Figure 2-figure supplement 3. Neural entrainment to syllable rate (4 Hz), excluding the final visit (n=54 recordings).

A. Design salience (left) and brain salience (right topography) derived from the significant latent component for neural entrainment to the syllable rate (4 Hz) using group (low versus high-likelihood for autism) as contrast. Significance (i.e., salience) was established through bootstrapping. Bars represent the mean of 500 random salience samples with replacement bootstrapping, and error bars indicate the 95% confidence interval. Yellow shading highlights variables that significantly contribute to the latent component, defined by a Bootstrap Ratio (BSR; mean of bootstrapping divided by standard deviation) > 2.3. The topography of the BSR values shows electrodes significantly contributing to the component (indicated by black dots, BSR > 2.3). **B.** Individual raw Phase Locking Values (PLVs) extracted from the salient electrodes identified by the latent component (black dots in the topography on 2A) are displayed. At every age, a verbal developmental quotient (DQ) of 100 is expected in the general population. A fitted curve is included for visualization purposes only, produced by a mixed-effects model with a 95% confidence interval. This curve is intended solely to aid visualization, as the statistical relationships between EEG and behavioral variables are determined by the PLS-C analysis.

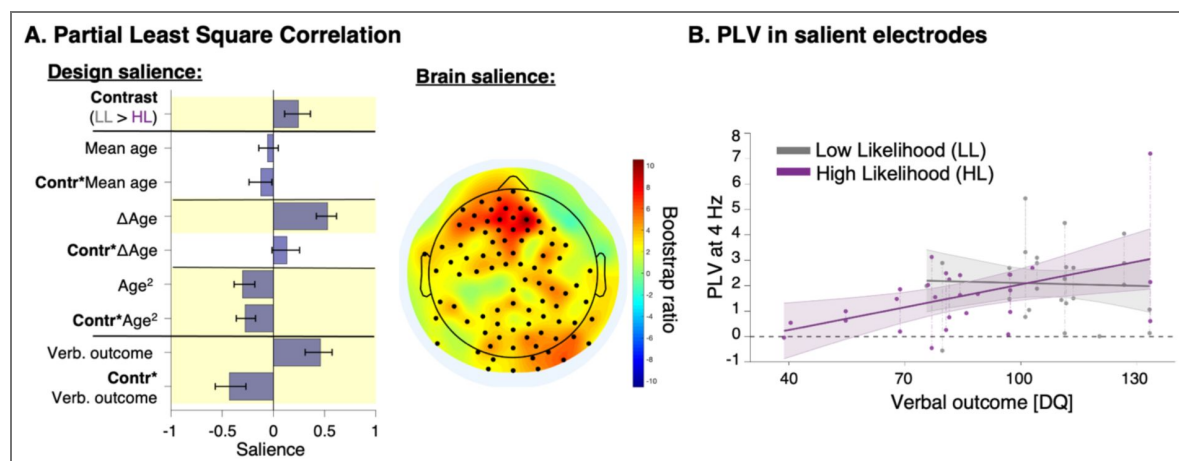


Figure 3-figure supplement 1. Neural entrainment to word rate in sleeping participants (n=25 recordings).

A. Design salience (left) and brain salience (right topography) derived from the significant latent component for neural entrainment to the word rate using frequency (1.3hz versus adjacent frequencies) as contrast. Significance (i.e., salience) was established through bootstrapping. **B.** Individual raw Phase Locking Values (PLVs) extracted from the salient electrodes identified by the latent component (black dots in the topography on A.) are displayed. A fitted curve is included for visualization purposes only, produced by a mixed-effects model with a 95% confidence interval.

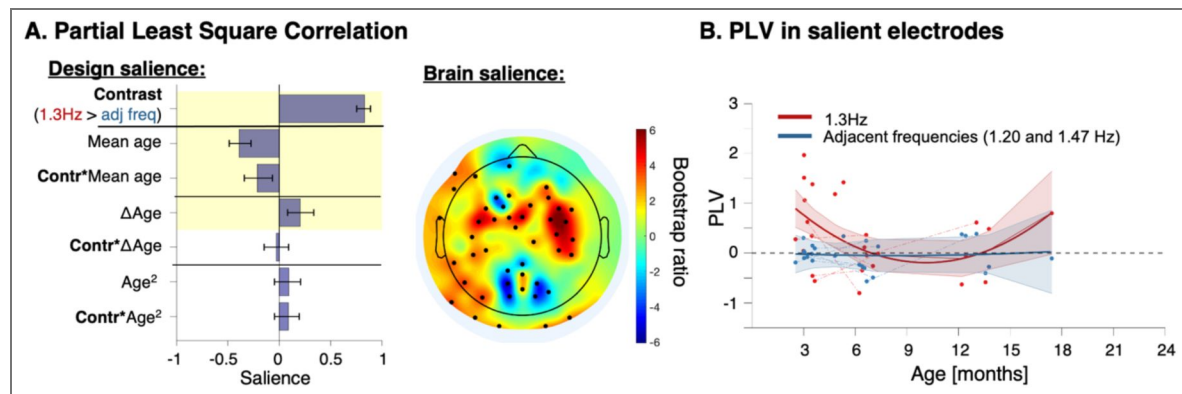
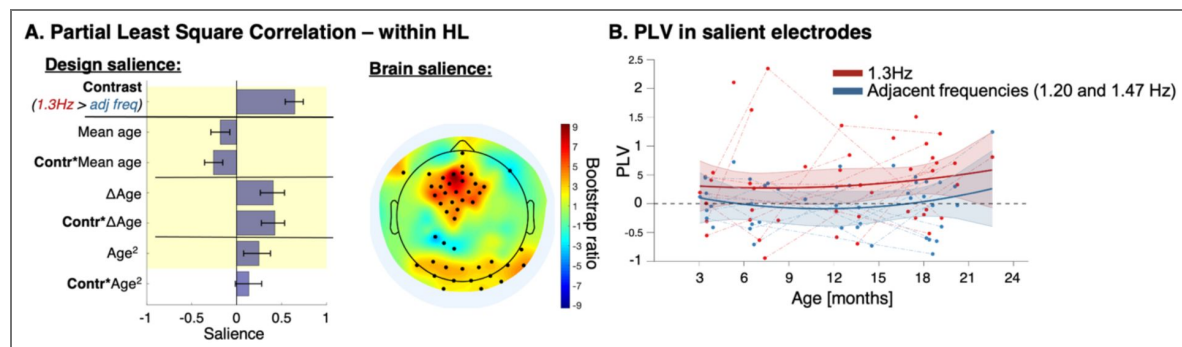


Figure 3-figure supplement 2. Word entrainment within HL participants (n=25; 44 recordings).

A. Design salience (left) and brain salience (right topography) derived from the significant latent component for neural entrainment to the word rate using frequency (1.3hz versus adjacent frequencies) as contrast. Significance (i.e., salience) was established through bootstrapping. BSRs: contrast : 12.9*; mean age: -3.2*; contrast*mean age: -5.3*; delta age: 5.9*; contrast*delta age: 6.7*; age²: 3.6*; contrast*age²: 1.9. **B.** Individual raw Phase Locking Values (PLVs) extracted from the salient electrodes identified by the latent component (black dots in the topography on A.) are displayed. A fitted curve is included for visualization purposes only, produced by a mixed-effects model with a 95% confidence interval.



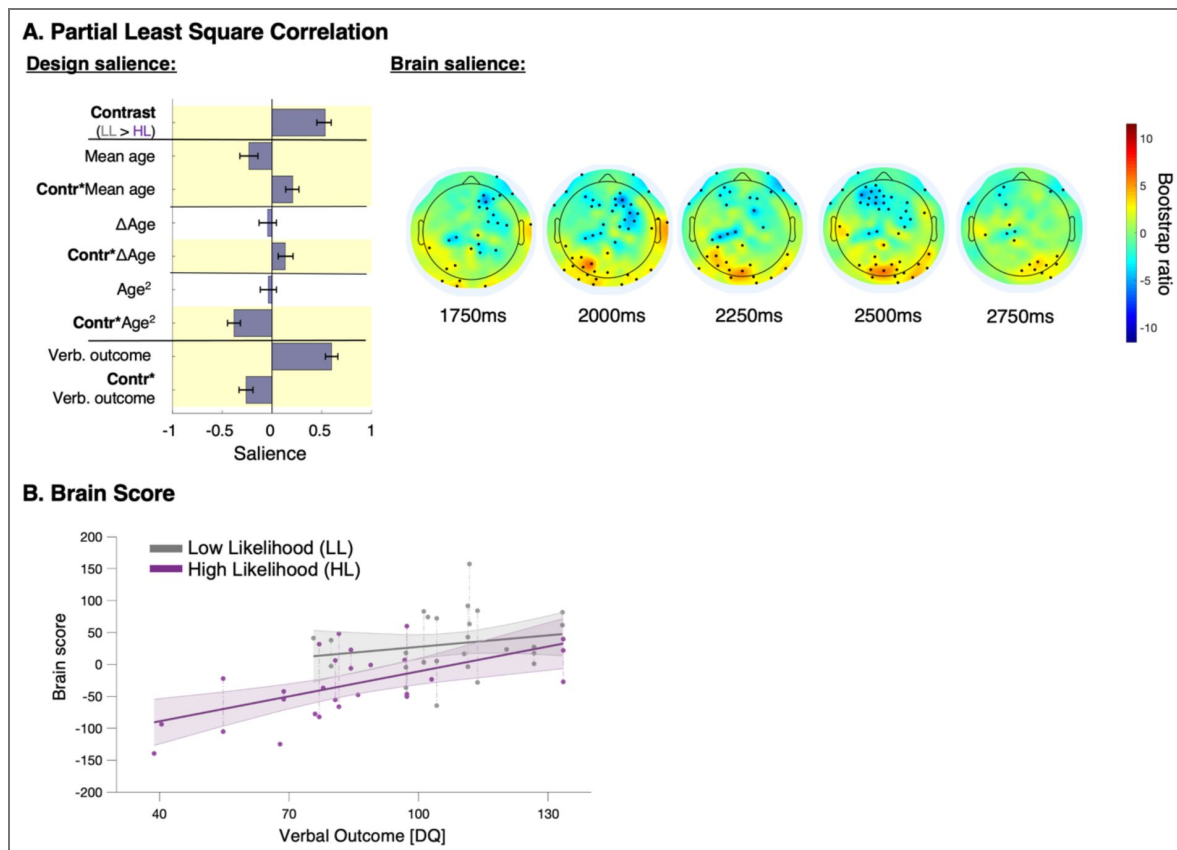


Figure 6 -figure supplement 1. Late evoked response potential (ERP) to word novelty, excluding the final visit (n=54 recordings).

A. Design and brain saliencies derived from the significant latent component. Brain topographies of bootstrap ratios (BSR) are displayed at 250ms intervals. Black dots indicate BSR > 2.3. **B.** Participants' brain scores for part-word and word conditions, as a function of verbal outcome. Brain scores are participants' raw voltage data projected onto electrode saliencies. Brain scores illustrate how individual EEG data fit the saliencies derived from the latent component. Linear fitting is used for illustrative purposes only. HL: high likelihood for autism; LL: low likelihood for autism.

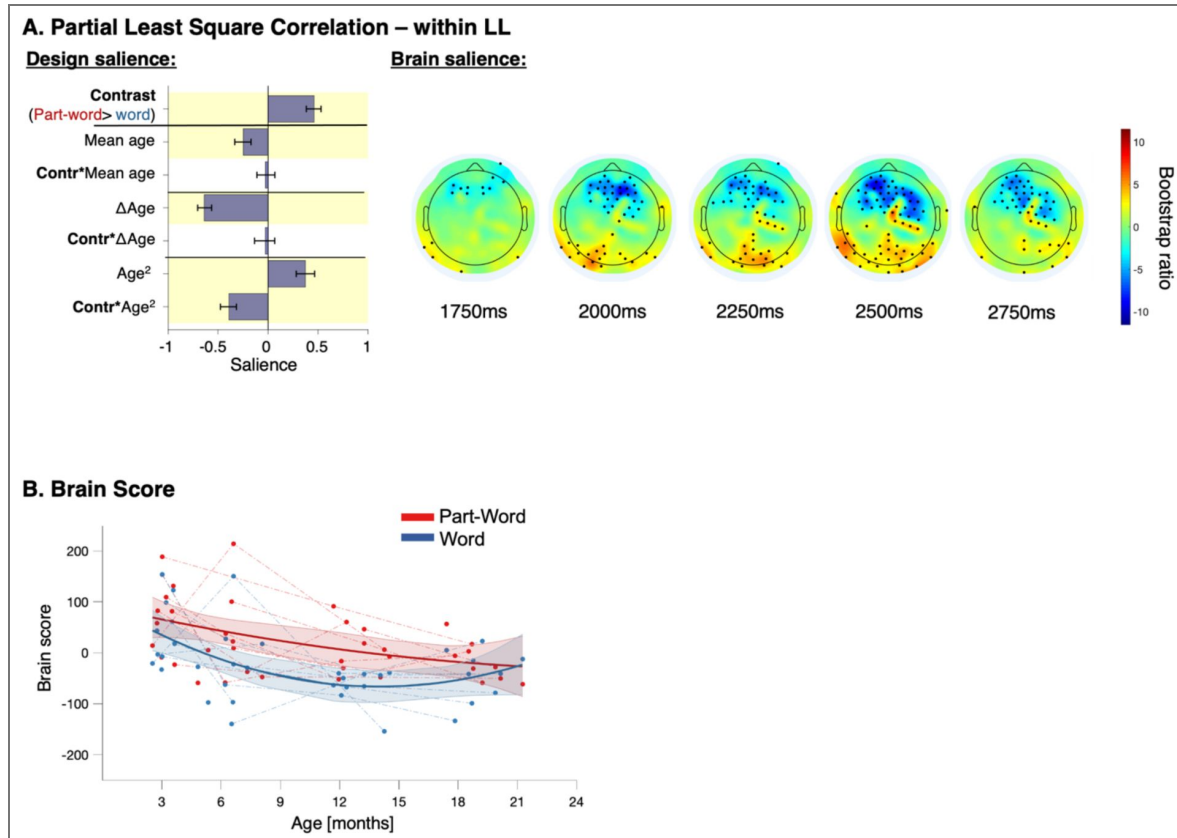


Figure 6 -figure supplement 2. Late evoked response potential (ERP) to word novelty within low likelihood infants (n=19; 39 recordings)

A. Design and brain saliences derived from the significant latent component. Brain topographies of bootstrap ratios (BSR) are displayed at 250ms intervals. Black dots indicate BSR > 2.3. **B.** Participants' brain scores for part-word and word conditions, as a function of age. Brain scores are participants' raw voltage data projected onto electrode saliences. Brain scores illustrate how individual EEG data fit the saliences derived from the latent component. Linear fitting is used for illustrative purposes only. LL: low likelihood for autism.

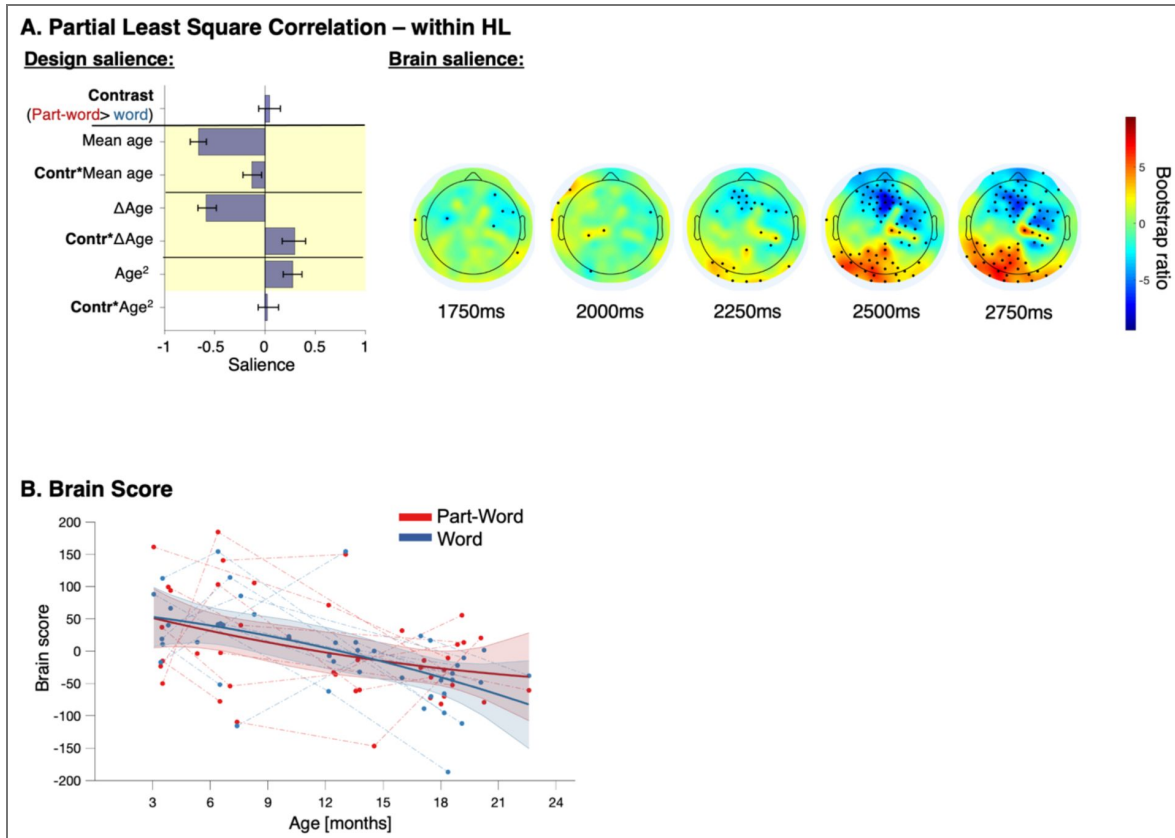


Figure 6 -figure supplement 3. Late evoked response potential (ERP) to word novelty within high likelihood infants (n=25; 44 recordings)

A. Design and brain saliences derived from the significant latent component. Brain topographies of bootstrap ratios (BSR) are displayed at 250ms intervals. Black dots indicate BSR > 2.3. **B.** Participants' brain scores for part-word and word conditions, as a function of age. Brain scores are participants' raw voltage data projected onto electrode saliences. Brain scores illustrate how individual EEG data fit the saliences derived from the latent component. Linear fitting is used for illustrative purposes only. HL: high likelihood for autism.

Supplementary Material

1) Stimuli

Using the MBROLA diphone, we synthesized 12 syllables (250 ms duration) using French phonemes: 6 vowels with 160ms duration ([i], [e], [ɛ], [a], [u], [o]) and 12 consonants with 90ms duration, 6 voiced ([b], [d], [g], [v], [z], [ʒ]), 6 unvoiced ([p], [t], [k], [f], [s], [ʃ]). Syllables were [ta], [do], [vɛ], [fi], [za], [pu], [ge], [kɛ], [so], [ʒu], [fe], and [bi]. All syllable combinations followed French phonotactic rules. Each syllable was produced by six selected speakers in MBROLA: 3 male adult speakers (fr3 with low pitch, fr1 with middle pitch, fr7 with high pitch) and 3 female adult speakers (fr2 with low pitch, it4 with middle pitch and fr4 with high pitch). Random (RND) and structured (STR) streams were constructed by concatenating the syllables without coarticulation and a random choice of the speaker, with the only constraints that there was no repetition of the same speaker in a row, nor alternation between two speakers more than once (if A and B are two speakers, then ABAB is forbidden). The same rule was applied for syllable identity in the RND stream and the word in the STR stream (i.e. no repetition, nor alternation more than once of a syllable or a word). Three STRs were built, each one using different words, and randomized across participants. The words in stream 1 were [tadovɛ], [fizapu], [gekeso], and [ʒufebi], and part-words (including the 2 last syllables of a word with a random first syllable of another word) were [dovɛfi], [ʒebita], [bitado], and [soʒufe]. In the second stream words were [dovɛfi], [zapuge], [kesoʒu], and [ʒebita]. Part-words were [vɛfike], [soʒufe], [tadovɛ], and [gekeso]. In the third stream, words were [vɛfike], [soʒufe], [pugeza], and [bitado]. Part-words were [tadovɛ], [ʒufebi], [kesoʒu], and [zapuge].

2) Data quality

A) Data quality for neural entrainment measures

We applied a linear mixed-effect model (MyMixedModelsTrajectories toolbox in Matlab R2018: <https://github.com/danizoeller/myMixedModelsTrajectories>) on individuals' amount of included epochs in RS, RND and STR (total of included epochs, the maximum possible value being 96), with fixed effects for age and group (HL versus LL), and random slope that varied by participant. The number of included epochs for each participant was used as a proxy for data quality. Three random-slope mixed-effect models with random slopes (constant, linear, quadratic) were fitted using the following formula:

(1) Good-epochs $\sim$ age * group + (1+age | subject)

The following equation was used for the linear model:

(2) Good-epochs_{im} = $\beta_0 + \beta_1 \text{group}_i + \beta_2 \text{age}_{im} + \beta_3 \text{group}_i * \text{age}_{im} + b_{1m} * \text{age}_i + b_{0m} + \epsilon_{im}$

for participant i at timepoint m , with β_1 -3 being the fixed-effect coefficients for group (coded by a dummy variable), age and group*age interaction. The b_{1m} term is the random slope varying by participant, b_{0m} is the normally distributed random effect, and ϵ_{im} is the normally distributed observation error. The Bayesian information criterion (BIC) indicated the best-fitting model. Here, the constant model was selected (BIC=584), suggesting that age had no effect on data quality. We found no significant group effect ($p=.748$, $\beta=85.5$). Consequently, we considered that data quality for neural entrainment measures was broadly similar between HL and LL and across ages. This analysis is illustrated in the [Figure S1](#).

B) ERP data quality

The same analysis was run on the ERP epochs (total of included epochs, the largest possible value being 96: 48 Word items and 48 Part-Word items). The model of order 1 (linear) was selected (BIC=569), we found no significant group effect ($p=.656$, $\beta=-1.6$) or group*age interaction effect ($p=.436$, $\beta=-0.4$) or age effect ($p=.149$, $\beta=-0.6$). This analysis is illustrated on the [Figure S2](#).

Figure S1. Neural entrainment data quality.

LL mean trajectory is in red and HL in blue, with their respective 95% confidence intervals. Included epochs in absolute number.

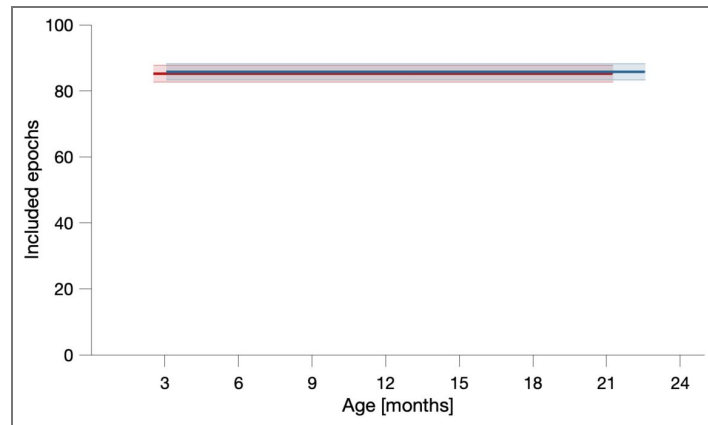
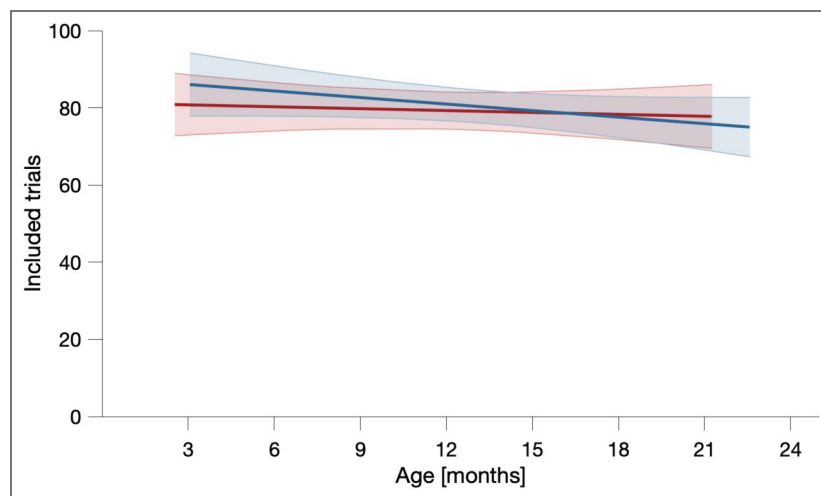


Figure S2. ERP data quality.

LL mean trajectory is in red and HL in blue, with their respective 95% confidence intervals.



3) Supplementary analyses for neural entrainment

Learning time course across all participants

Additionally, we ran an analysis of neural entrainment over the time course of the stimuli presentation to provide an illustration of neural entrainment over the experimental session, as done in Smalle et al (2022) and Flò et al (2022) for instance. We concatenated the epochs in chronological order (180 seconds of random stream, 360 of structured stream, and 180s of random stream again). For each targeted frequency (word at 1.33 Hz and syllable at 4 Hz), we averaged PLVs over the salient electrodes obtained from the PLS-C (black dots from Figure 3C for 1.3Hz entrainment and from Figure 2C for 4Hz). For both frequencies, we fitted a mixed-effect model using Matlab *fitlme* function ($PLV \sim 1 + (1 | \text{participant})$), using 120 second sliding windows with an increment of 1.5 second and testing for differences with zero (Figure S3).

Group differences in the time course of neural entrainment

We ran two PLS-cs to analyze this time dimension. Brain matrices included PLVs at each timeframe (120s time-windows with 1.5s increment) averaged in space across salient electrodes obtained from the PLS in space dimension (black dots on Figure 3C for word entrainment and black dots of Figure 2C for syllable entrainment). Design matrices were the same as in previous analyses (including group LL>HL as contrast and age-related variables and verbal outcome as factors).

The PLS-c for 4Hz syllable entrainment resulted in one significant latent component ($p=.001$; $r=.66$; 32.0% explained covariance) with the following BSRs: contrast (LL > HL): 7.2*; mean age: -12.3*; contrast*mean age: -1.4; delta age: 5.9*; contrast*delta age: -2.7*; age²: -1.8; contrast*age²: 2.2; verbal outcome: 20.5*; contrast*verbal outcome: -12.8* (Figure S4). PLS-c for 1.3Hz word entrainment gave no significant latent component (most significant latent component : $p=.103$) revealing no difference between HL and LL participants.

To better disentangle how each group tracks syllables in STR in contrast to RND, we ran a PLS-c testing group effect on 4hz PLVs in each stream. In the RND stream, the analysis yields one significant component ($p<.001$, $r=.49$, 52.7% explained covariance, Figure S5A-B). Bootstrap ratios (BSR) were: contrast (low versus high likelihood) 6.9; mean age -1.0; contrast*mean-age 0.1; delta-age 11.8; contrast*delta-age 0.1; age² -4.9; contrast*age² 4.6; verbal-outcome 9.7; contrast*verbal-outcome -0.6. The model still highlighted a strong link between syllable tracking and group, suggesting that RND also discriminates between HL and LL. However, RND syllable tracking doesn't appear to be linked to group x verbal-outcome as we observed in Figure 2C-D). In the STR stream, we obtained one significant latent component ($p=.002$, $r=.51$, 57.8% explained covariance, Figure S5C-D). Bootstrap ratios (BSR) are: contrast 2.2; mean age -1.6; contrast*mean-age -1.1; delta-age 6.1; contrast*delta-age -0.7; age² -4.4; contrast*age² 0.5; verbal-outcome 8.2; contrast*verbal-outcome -7.3. Here, the strong association between syllable tracking and group*verbal-outcome is similar to the model presented in Figure 2C-D).

4) Supplementary analyses for ERP

First, ERP topographies were inspected across ages in the whole sample to exclude the presence of maturational qualitative changes that would have prevented the inclusion of all recordings into the same longitudinal PLSC model. Figure S6 shows the ERP topographies averaged across all participants at each visit for each condition (Part-word and Word), as well as the difference between conditions (Part-Word minus Word). No salient maturational qualitative change emerges from visual inspection (e.g., polarity inversion), legitimizing the use of a single multivariate linear model (PLS-C) that includes all visits.

Figures S7-S8 further divide ERP topographies by group, showing the response in both LL and HL at each age bin and for each condition: 3 months ($n=18$), 6-9 months ($n=20$), 12-15 months ($n=20$), 18-21 months ($n=25$).

Figure S3. Neural entrainment time course over the experimental session considering all participants.

The plain squares under the plots correspond to the time samples with PLVs significantly greater than 0 ($p < .05$).

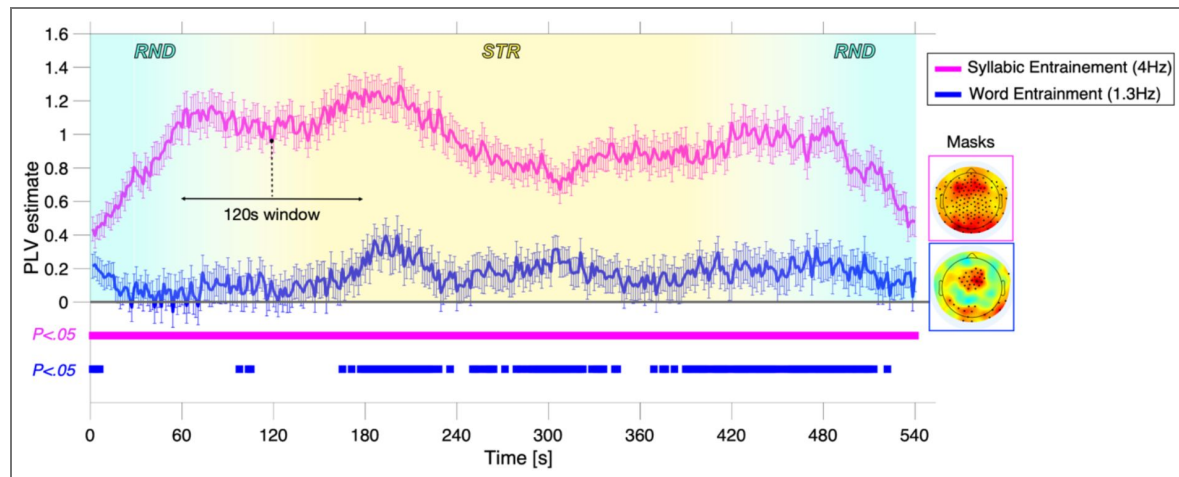


Figure S4. Group differences in the time course of syllable neural entrainment (4Hz).

Gray shading on right panel indicates $BSR < 2.3$; $BSR > 2.3$ is considered significant. Vertical dashed lines indicate the transitions between random and structured streams.

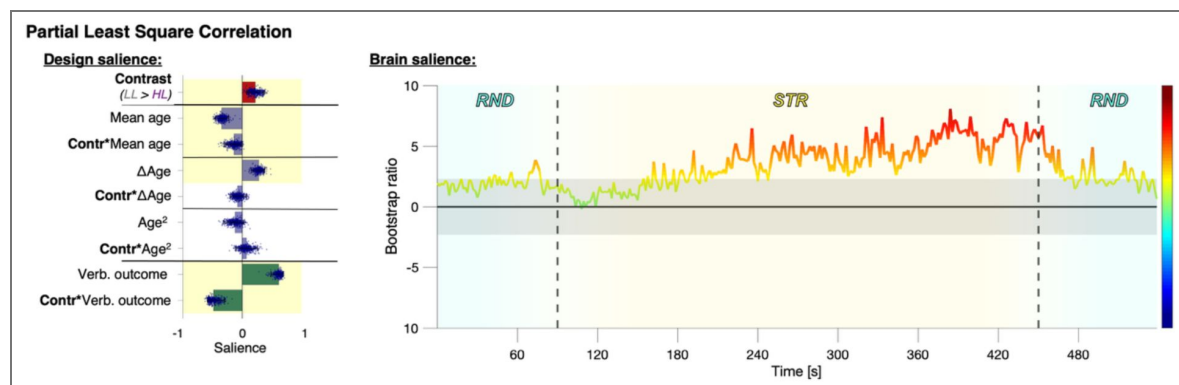


Figure S5. Syllable entrainment within RND (A-B) and STR (C-D).

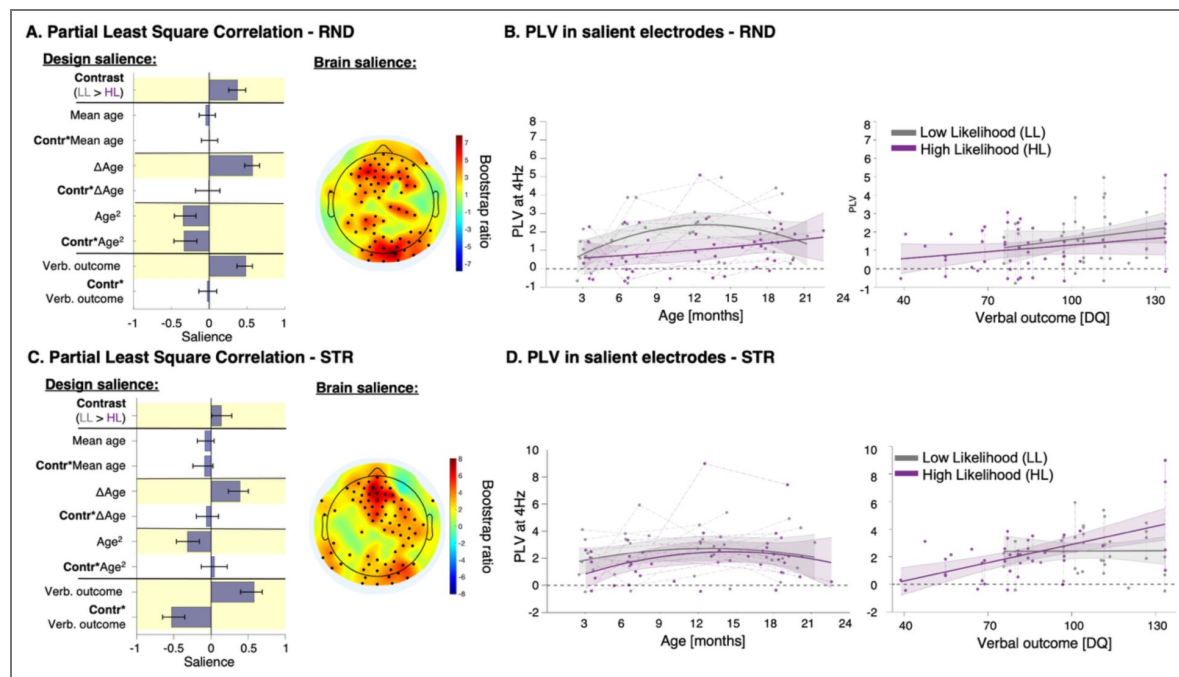
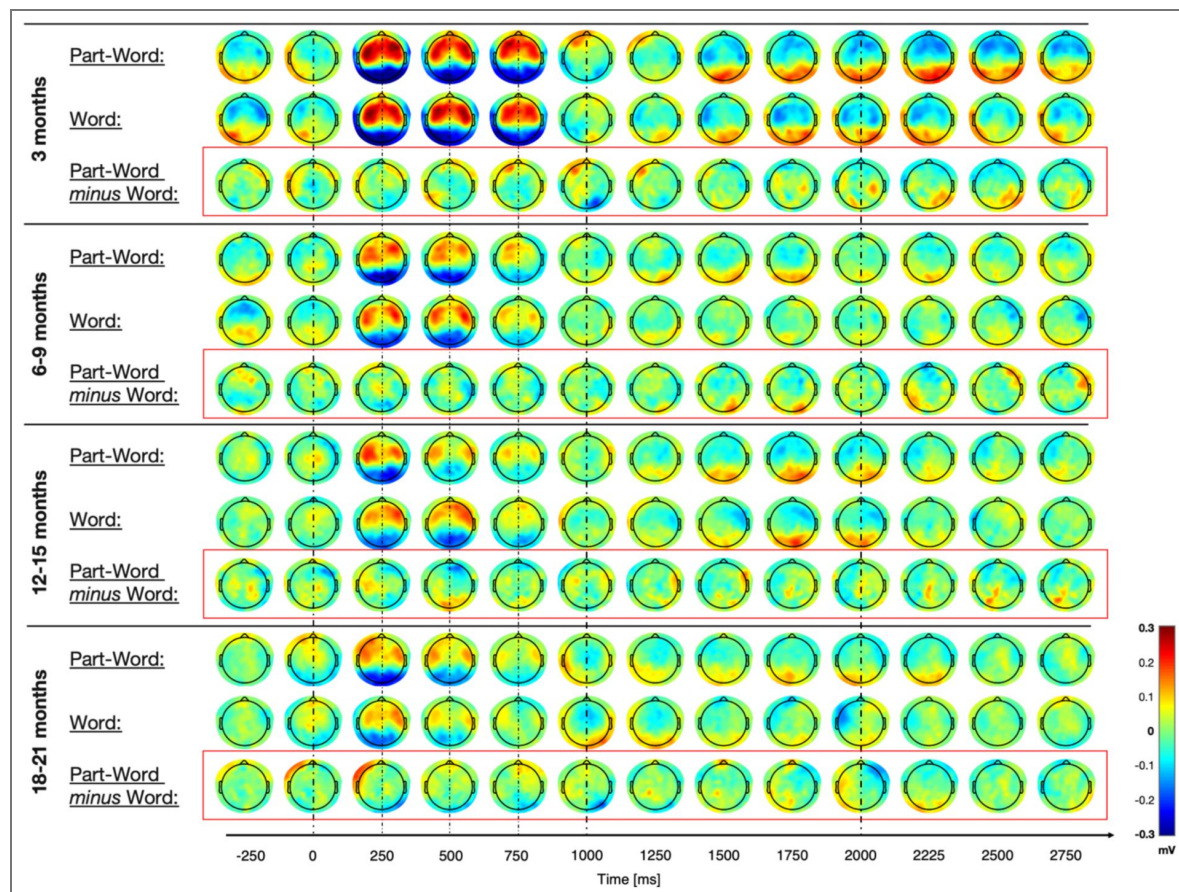


Figure S6. ERP topographies across age bins and participants



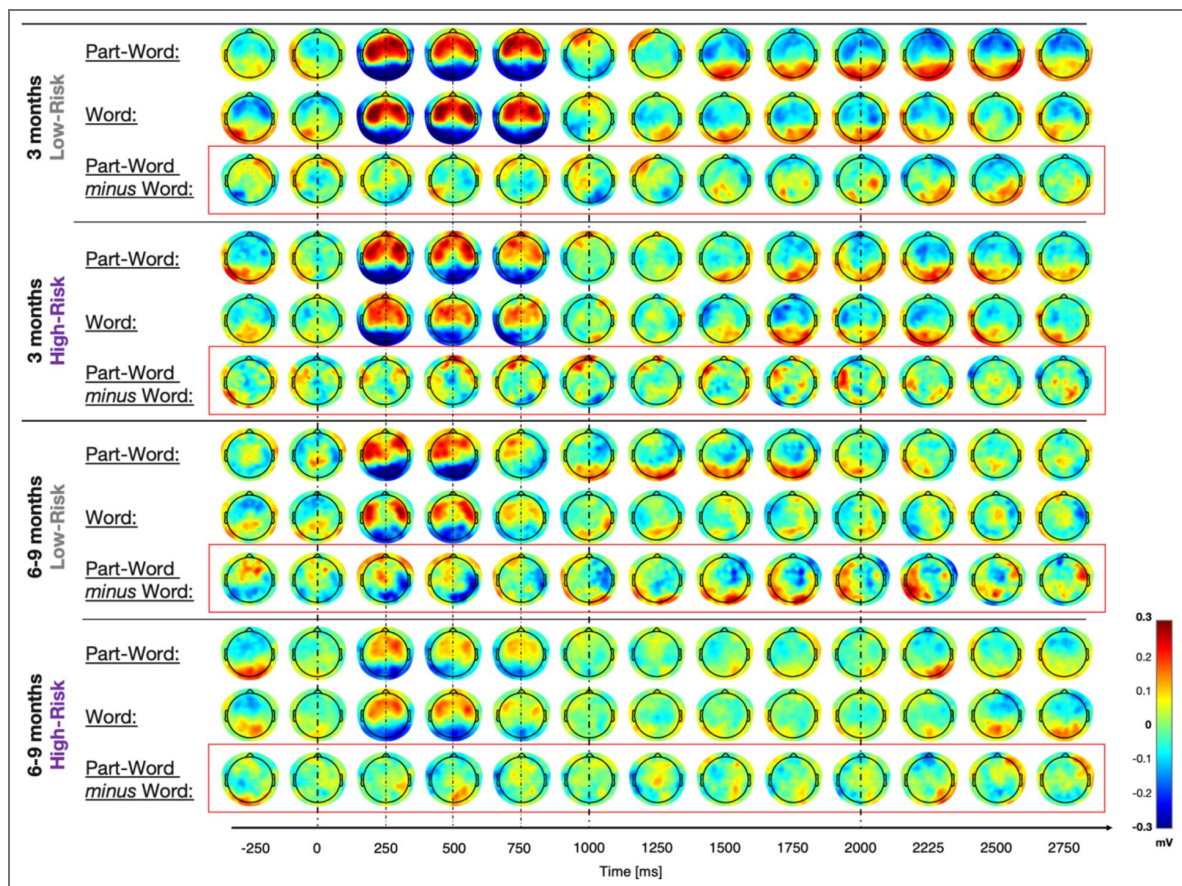


Figure S7. ERP topographies in each group at 3mo and 6-9mo

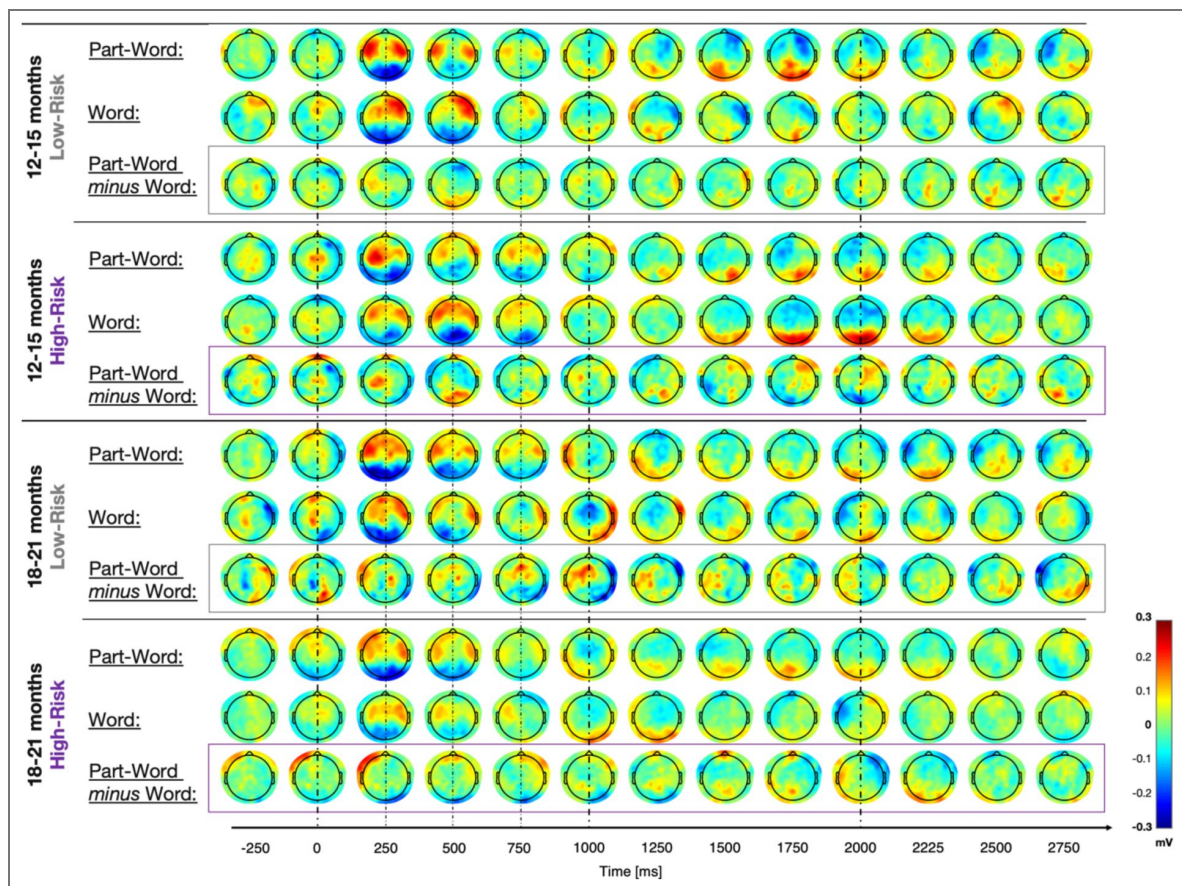


Figure S8. ERP topographies in each group at 12-15mo and 18-21mo

Data availability

The corresponding author will make available the anonymised data used to support the conclusion of this work on reasonable request.

Acknowledgements

The authors would like to thank all the families who participated in the study, as well as the many collaborators who contributed to data collection over the years, especially Lylia Ben Hadid, Nada Kojovic, Kenza Latrèche, Sara Maglio, Irène Pittet, and Stefania Solazzo. We also would like to thank Farnaz Delavari for the invaluable pieces of advice provided for the statistical analyses.

Additional information

Contribution

A.F., G.D.L., M.G. and M.S. conceptualized the experimental protocol and analytic strategy. M.G. acquired and preprocessed the data. M.G. analyzed the data under the supervision of A.F. M.G. wrote the manuscript with the input of all authors. All authors read and approved the final manuscript.

Funding

This research was supported by the Swiss National Foundation Synapsy (Grant No. 51NF40–185897), the Swiss National Foundation for Scientific Research (Grant Nos. #191227 to M.G. and #163859, #190084, #202235, #212653 to M.S.), the Fondation Privée des Hôpitaux Universitaires de Genève (<https://www.fondationhug.org>) and by the Fondation Pôle Autisme (<https://www.pole-autisme.ch>). The funders were not involved in this study and had no role other than to provide financial support.

Funding

Funder	Grant reference number	Author
Schweizerischer Nationalfonds zur Förderung der Wissenschaftlichen Forschung (SNF)	51NF40-185897	Marie Schaer
Schweizerischer Nationalfonds zur Förderung der Wissenschaftlichen Forschung (SNF)	#191227	Michel Godel
Schweizerischer Nationalfonds zur Förderung der Wissenschaftlichen Forschung (SNF)	#163859	Michel Godel
Schweizerischer Nationalfonds zur Förderung der Wissenschaftlichen Forschung (SNF)	#190084	Marie Schaer
Schweizerischer Nationalfonds zur Förderung der Wissenschaftlichen Forschung (SNF)	#202235	Marie Schaer
Schweizerischer Nationalfonds zur Förderung der Wissenschaftlichen Forschung (SNF)	#212653	Marie Schaer
Fondation privée des Hôpitaux universitaires de Genève (HUG)		Marie Schaer
Fondation Pôle Autisme (FPA)		Marie Schaer

Author ORCID iDs

Michel Godel:  <https://orcid.org/0000-0002-9789-3540>

References

1. **American Psychiatric Association** (2013) Diagnostic and Statistical Manual of Mental Disorders. *American Psychiatric Association* <https://doi.org/10.1176/appi.books.9780890425596>
2. **Shaw KA**, Williams S, Patrick ME, Valencia-Prado M, Durkin MS, Howerton EM, et al. (2025) Prevalence and Early Identification of Autism Spectrum Disorder Among Children Aged 4 and 8 Years — Autism and Developmental Disabilities Monitoring Network, 16 Sites, United States, 2022. *MMWR Surveill Summ* **74**:1-22 <https://doi.org/10.15585/mmwr.ss7402a1> | [PubMed](#)
3. **Taylor MJ**, Rosenqvist MA, Larsson H, Gillberg C, D’Onofrio BM, Lichtenstein P, et al. (2020) Etiology of Autism Spectrum Disorders and Autistic Traits Over Time. *JAMA Psychiatry* **77**:936 <https://doi.org/10.1001/jamapsychiatry.2020.0680> | [PubMed](#)
4. **Latrèche K**, Godel M, Franchini M, Journal F, Kojovic N, Schaer M (2024) Early trajectories and moderators of autistic language profiles: A longitudinal study in preschoolers. *Autism* <https://doi.org/10.1177/13623613241253015> | [PubMed](#)
5. **Schaeffer J**, Abd El-Raziq M, Castroviejo E, Durrleman S, Ferré S, Grama I, et al. (2023) Language in autism: domains, profiles and co-occurring conditions. *J Neural Transm* **130**:433-57 <https://doi.org/10.1007/s00702-023-02592-y> | [PubMed](#)
6. **Hudry K**, Chandler S, Bedford R, Pasco G, Gliga T, Elsabbagh M, et al. (2014) Early Language Profiles in Infants at High-Risk for Autism Spectrum Disorders. *J Autism Dev Disord* **44**:154-67 <https://doi.org/10.1007/s10803-013-1861-4> | [PubMed](#)
7. **Lazenby DC**, Sideridis GD, Huntington N, Prante M, Dale PS, Curtin S, et al. (2016) Language Differences at 12 Months in Infants Who Develop Autism Spectrum Disorder. *J Autism Dev Disord* **46**:899-909 <https://doi.org/10.1007/s10803-015-2632-1> | [PubMed](#)
8. **Schreibman L**, Dawson G, Stahmer AC, Landa R, Rogers SJ, McGee GG, et al. (2015) Naturalistic Developmental Behavioral Interventions: Empirically Validated Treatments for Autism Spectrum Disorder. *J Autism Dev Disord* **45**:2411-28 <https://doi.org/10.1007/s10803-015-2407-8> | [PubMed](#)
9. **Miranda A**, Berenguer C, Baixauli I, Roselló B (2023) Childhood language skills as predictors of social, adaptive and behavior outcomes of adolescents with autism spectrum disorder. *Research in Autism Spectrum Disorders* **103** <https://doi.org/10.1016/j.rasd.2023.102143>
10. **Whitehouse AJO**, Varcin KJ, Pillar S, Billingham W, Alvares GA, Barbaro J, et al. (2021) Effect of Preemptive Intervention on Developmental Outcomes Among Infants Showing Early Signs of Autism: A Randomized Clinical Trial of Outcomes to Diagnosis. *JAMA Pediatr* **175**:e213298 <https://doi.org/10.1001/jamapediatrics.2021.3298> | [PubMed](#)
11. **Williams D** (1999) *Nobody nowhere: the remarkable autobiography of an autistic girl* London: Kingsley. pp. 192
12. **Mattys SL**, White L, Melhorn JF (2005) Integration of Multiple Speech Segmentation Cues: A Hierarchical Framework. *Journal of Experimental Psychology: General* **134**:477-500 <https://doi.org/10.1037/0096-3445.134.4.477> | [PubMed](#)
13. **Saffran JR**, Aslin RN, Newport EL (1996) Statistical Learning by 8-Month-Old Infants. *Science* **274**:1926-8 <https://doi.org/10.1126/science.274.5294.1926> | [PubMed](#)
14. **Fló A**, Brusini P, Macagno F, Nespor M, Mehler J, Ferry AL (2019) Newborns are sensitive to multiple cues for word segmentation in continuous speech. *Dev Sci* **22**:e12802 <https://doi.org/10.1111/desc.12802> | [PubMed](#)
15. **Fló A**, Benjamin L, Palu M, Dehaene-Lambertz G (2022) Sleeping neonates track transitional probabilities in speech but only retain the first syllable of words. *Sci Rep* **12**:4391 <https://doi.org/10.1038/s41598-022-08411-w> | [PubMed](#)
16. **Choi D**, Batterink LJ, Black AK, Paller KA, Werker JF (2020) Preverbal Infants Discover Statistical Word Patterns at Similar Rates as Adults: Evidence From Neural Entrainment. *Psychol Sci* **31**:1161-73 <https://doi.org/10.1177/0956797620933237> | [PubMed](#)

17. Bulf H, Johnson SP, Valenza E (2011) Visual statistical learning in the newborn infant. *Cognition* **121**:127-32 <https://doi.org/10.1016/j.cognition.2011.06.010> | PubMed
18. Fiser J, Aslin RN (2002) Statistical learning of new visual feature combinations by infants. *Proc Natl Acad Sci USA* **99**:15822-6 <https://doi.org/10.1073/pnas.232472899> | PubMed
19. Conway CM, Christiansen MH (2005) Modality-Constrained Statistical Learning of Tactile, Visual, and Auditory Sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition* **31**:24-39 <https://doi.org/10.1037/0278-7393.31.1.24> | PubMed
20. Benjamin L, Zang D, Fló A, Qi Z, Su P, Zhou W, et al. (2025) The role of conscious attention in auditory statistical learning: Evidence from patients with impaired consciousness. *iScience* **28**:111591 <https://doi.org/10.1016/j.isci.2024.111591> | PubMed
21. Evans JL, Saffran JR, Robe-Torres K (2009) Statistical Learning in Children With Specific Language Impairment. *J Speech Lang Hear Res* **52**:321-35 [https://doi.org/10.1044/1092-4388\(2009/07-0189\)](https://doi.org/10.1044/1092-4388(2009/07-0189)) | PubMed
22. Lammertink I, Boersma P, Wijnen F, Rispens J (2017) Statistical Learning in Specific Language Impairment: A Meta-Analysis. *J Speech Lang Hear Res* **60**:3474-86 https://doi.org/10.1044/2017_JSLHR-L-16-0439 | PubMed
23. Saffran JR (2018) Statistical learning as a window into developmental disabilities. *J Neurodevelopmental Disord* **10**:35 <https://doi.org/10.1186/s11689-018-9252-y> | PubMed
24. Gabay Y, Thiessen ED, Holt LL (2015) Impaired Statistical Learning in Developmental Dyslexia. *J Speech Lang Hear Res* **58**:934-45 https://doi.org/10.1044/2015_JSLHR-L-14-0324 | PubMed
25. Szatmari P, Chawarska K, Dawson G, Georgiades S, Landa R, Lord C, et al. (2016) Prospective Longitudinal Studies of Infant Siblings of Children With Autism: Lessons Learned and Future Directions. *J Am Acad Child Adolesc Psychiatry* **55**:179-87 <https://doi.org/10.1016/j.jaac.2015.12.014> | PubMed | PubMed Central
26. Wilkinson CL, Gabard-Durnam LJ, Kapur K, Tager-Flusberg H, Levin AR, Nelson CA (2020) Use of Longitudinal EEG Measures in Estimating Language Development in Infants With and Without Familial Risk for Autism Spectrum Disorder. *Neurobiology of Language* **1**:33-53 https://doi.org/10.1162/nol_a_00002 | PubMed
27. McDonald NM, Jeste SS (2021) Beyond Baby Siblings—Expanding the Definition of “High-Risk Infants” in Autism Research. *Curr Psychiatry Rep* **23**:34 <https://doi.org/10.1007/s11920-021-01243-x> | PubMed
28. Jochaut D, Lehongre K, Saitovitch A, Devauchelle AD, Olasagasti I, Chabane N, et al. (2015) Atypical coordination of cortical oscillations in response to speech in autism. *Front Hum Neurosci* **9** <https://doi.org/10.3389/fnhum.2015.00171> | PubMed
29. Menn KH, Ward EK, Braukmann R, Van Den Boomen C, Buitelaar J, Hunnius S, et al. (2022) Neural Tracking in Infancy Predicts Language Development in Children With and Without Family History of Autism. *Neurobiology of Language* **3**:495-514 https://doi.org/10.1162/nol_a_00074 | PubMed
30. Wang X, Delgado J, Marchesotti S, Kojovic N, Sperdin HF, Rihs TA, et al. (2023) Speech Reception in Young Children with Autism Is Selectively Indexed by a Neural Oscillation Coupling Anomaly. *J Neurosci* **43**:6779-95 <https://doi.org/10.1523/jneurosci.0112-22.2023> | PubMed
31. Edgar EV, McGuire K, Pelphrey KA, Ventola P, Van Noordt S, Crowley MJ (2024) Early- and Late-Stage Auditory Processing of Speech Versus Non-Speech Sounds in Children With Autism Spectrum Disorder: An ERP and Oscillatory Activity Study. *Developmental Psychobiology* **66** <https://doi.org/10.1002/dev.22552> | PubMed
32. Bruining H, Hardstone R, Juarez-Martinez EL, Sprengers J, Avramiea AE, Simpraga S, et al. (2020) Measurement of excitation-inhibition ratio in autism spectrum disorder using critical brain dynamics. *Sci Rep* **10**:9195 <https://doi.org/10.1038/s41598-020-65500-4> | PubMed

33. Pagano J, Landi S, Stefanoni A, Nardi G, Albanesi M, Bauer HF, et al. (2023) Shank3 deletion in PV neurons is associated with abnormal behaviors and neuronal functions that are rescued by increasing GABAergic signaling. *Mol Autism* **14**:28 <https://doi.org/10.1186/s13229-023-00557-2> | PubMed
34. Hollestein V, Poelmans G, Forde NJ, Beckmann CF, Ecker C, Mann C, et al. (2023) Excitatory/inhibitory imbalance in autism: the role of glutamate and GABA gene-sets in symptoms and cortical brain structure. *Transl Psychiatry* **13**:18 <https://doi.org/10.1038/s41398-023-02317-5> | PubMed
35. Russo N, Zecker S, Trommer B, Chen J, Kraus N (2009) Effects of Background Noise on Cortical Encoding of Speech in Autism Spectrum Disorders. *J Autism Dev Disord* **39**:1185-96 <https://doi.org/10.1007/s10803-009-0737-0> | PubMed
36. Song JH, Banai K, Russo NM, Kraus N (2006) On the Relationship between Speech- and Nonspeech-Evoked Auditory Brainstem Responses. *Audiol Neurotol* **11**:233-41 <https://doi.org/10.1159/000093058> | PubMed
37. Tecoulesco L, Skoe E, Naigles LR (2020) Phonetic discrimination mediates the relationship between auditory brainstem response stability and syntactic performance. *Brain and Language* **208** <https://doi.org/10.1016/j.bandl.2020.104810> | PubMed
38. Hu A, Kozloff V, Van Horne A, Owen, Chugani D, Qi Z (2024) Dissociation Between Linguistic and Nonlinguistic Statistical Learning in Children with Autism. *J Autism Dev Disord* **54**:1912-27 <https://doi.org/10.1007/s10803-023-05902-1> | PubMed
39. Wagley N, Lajiness-O'Neill R, Hay JSF, Bowyer SM, Ugolini M, Kovelman I, et al. (2020) Predictive processing during a naturalistic statistical learning task in ASD. *eNeuro* ENEURO.0069-19.2020 <https://doi.org/10.1523/ENEURO.0069-19.2020> | PubMed
40. Scott-Van Zeeland AA, McNealy K, Wang AT, Sigman M, Bookheimer SY, Dapretto M (2010) No Neural Evidence of Statistical Learning During Exposure to Artificial Languages in Children with Autism Spectrum Disorders. *Biological Psychiatry* **68**:345-51 <https://doi.org/10.1016/j.biopsych.2010.01.011> | PubMed
41. Liu J, Tsang T, Ponting C, Jackson L, Jeste SS, Bookheimer SY, et al. (2021) Lack of Neural Evidence for Implicit Language Learning in 9-Month-Old Infants at High Risk for Autism. *Developmental Science* <https://doi.org/10.1111/desc.13078> | PubMed
42. Forest TA, Schlichting ML, Duncan KD, Finn AS (2023) Changes in statistical learning across development. *Nat Rev Psychol* **2**:205-19 <https://doi.org/10.1038/s44159-023-00157-0>
43. Benjamin L, Fló A, Palu M, Naik S, Melloni L, Dehaene-Lambertz G (2023) Tracking transitional probabilities and segmenting auditory sequences are dissociable processes in adults and neonates. *Developmental Science* **26** <https://doi.org/10.1111/desc.13300> | PubMed
44. Fló A, Benjamin L, Palu M, Dehaene-Lambertz G (2025) Statistical learning beyond words in human neonates. *eLife* **13**:RP101802 <https://doi.org/10.7554/eLife.101802> | PubMed
45. Buiatti M, Peña M, Dehaene-Lambertz G (2009) Investigating the neural correlates of continuous speech computation with frequency-tagged neuroelectric responses. *NeuroImage* **44**:509-19 <https://doi.org/10.1016/j.neuroimage.2008.09.015> | PubMed
46. Kabdebon C, Pena M, Buiatti M, Dehaene-Lambertz G (2015) Electrophysiological evidence of statistical learning of long-distance dependencies in 8-month-old preterm and full-term infants. *Brain and Language* **148**:25-36 <https://doi.org/10.1016/j.bandl.2015.03.005> | PubMed
47. Happé F, Frith U (2006) The Weak Coherence Account: Detail-focused Cognitive Style in Autism Spectrum Disorders. *J Autism Dev Disord* **36**:5-25 <https://doi.org/10.1007/s10803-005-0039-0> | PubMed
48. Park WJ, Schauder KB, Zhang R, Bennetto L, Tadin D (2017) High internal noise and poor external noise filtering characterize perception in autism spectrum disorder. *Sci Rep* **7**:17584 <https://doi.org/10.1038/s41598-017-17676-5> | PubMed

49. Gennari G, Marti S, Palu M, Fló A, Dehaene-Lambertz G (2021) Orthogonal neural codes for speech in the infant brain. *Proc Natl Acad Sci USA* **118**:e2020410118 <https://doi.org/10.1073/pnas.2020410118> | PubMed
50. Krishnan A, Williams LJ, McIntosh AR, Abdi H (2011) Partial Least Squares (PLS) methods for neuroimaging: A tutorial and review. *NeuroImage* **56**:455-75 <https://doi.org/10.1016/j.neuroimage.2010.07.034> | PubMed
51. Delavari F, Sandini C, Zöller D, Mancini V, Bortolin K, Schneider M, et al. (2021) Dysmaturation Observed as Altered Hippocampal Functional Connectivity at Rest Is Associated With the Emergence of Positive Psychotic Symptoms in Patients With 22q11 Deletion Syndrome. *Biological Psychiatry* **90**:58-68 <https://doi.org/10.1016/j.biopsych.2020.12.033> | PubMed
52. Peng H, Lu Y (2012) Model selection in linear mixed effect models. *Journal of Multivariate Analysis* **109**:109-29 <https://doi.org/10.1016/j.jmva.2012.02.005>
53. Csibra G, Kushnirenko E, Grossmann T (2008) Electrophysiological Methods in Studying Infant Cognitive Development. In: Nelson CA, Luciana M (Eds). *Handbook of Developmental Cognitive Neuroscience* (2nd ed) Cambridge, MA: MIT Press. pp. 247-62
54. Giraud AL, Poeppel D (2012) Cortical oscillations and speech processing: emerging computational principles and operations. *Nat Neurosci* **15**:511-7 <https://doi.org/10.1038/nn.3063> | PubMed
55. Russo N, Nicol T, Trommer B, Zecker S, Kraus N (2009) Brainstem transcription of speech is disrupted in children with autism spectrum disorders. *Developmental Science* **12**:557-67 <https://doi.org/10.1111/j.1467-7687.2008.00790.x> | PubMed
56. Lehongre K, Ramus F, Villiermet N, Schwartz D, Giraud AL (2011) Altered Low-Gamma Sampling in Auditory Cortex Accounts for the Three Main Facets of Dyslexia. *Neuron* **72**:1080-90 <https://doi.org/10.1016/j.neuron.2011.11.002> | PubMed
57. Seymour RA, Rippon G, Gooding-Williams G, Sowman PF, Kessler K (2020) Reduced auditory steady state responses in autism spectrum disorder. *Molecular Autism* **11**:56 <https://doi.org/10.1186/s13229-020-00357-y> | PubMed
58. Patel SP, Winston M, Guilfoyle J, Nicol T, Martin GE, Nayar K, et al. (2023) Neural Processing of Speech Sounds in ASD and First-Degree Relatives. *J Autism Dev Disord* **53**:3257-71 <https://doi.org/10.1007/s10803-022-05562-7> | PubMed
59. Houston DM, Jusczyk PW (2000) The role of talker-specific information in word segmentation by infants. *Journal of Experimental Psychology: Human Perception and Performance* **26**:1570-82 <https://doi.org/10.1037/0096-1523.26.5.1570>
60. Harford EE, Holt LL, Abel TJ (2024) Unveiling the development of human voice perception: Neurobiological mechanisms and pathophysiology. *Current Research in Neurobiology* **6** <https://doi.org/10.1016/j.crneur.2024.100127> | PubMed
61. Shukla M, White KS, Aslin RN (2011) Prosody guides the rapid mapping of auditory word forms onto visual objects in 6-mo-old infants. *Proceedings of the National Academy of Sciences* **108**:6038-43 <https://doi.org/10.1073/pnas.1017617108> | PubMed
62. Choi JY, Kou RSN, Perrachione TK (2022) Distinct mechanisms for talker adaptation operate in parallel on different timescales. *Psychon Bull Rev* **29**:627-34 <https://doi.org/10.3758/s13423-021-02019-3> | PubMed
63. Luthra S (2024) Why are listeners hindered by talker variability?. *Psychon Bull Rev* **31**:104-21 <https://doi.org/10.3758/s13423-023-02355-6> | PubMed
64. Georgiades S, Szatmari P, Zwaigenbaum L, Bryson S, Brian J, Roberts W, et al. (2013) A Prospective Study of Autistic-Like Traits in Unaffected Siblings of Probands With Autism Spectrum Disorder. *JAMA Psychiatry* **70**:42 <https://doi.org/10.1001/2013.jamapsychiatry.1> | PubMed
65. Desaunay P, Briant AR, Bowler DM, Ring M, Gérardin P, Baleyte JM, et al. (2020) Memory in autism spectrum disorder: A meta-analysis of experimental studies. *Psychological Bulletin* **146**:377-410 <https://doi.org/10.1037/bul0000225> | PubMed

66. Vivanti G, Hocking DR, Fanning PAJ, Uljarevic M, Postorino V, Mazzone L, et al. (2018) Attention to novelty versus repetition: Contrasting habituation profiles in Autism and Williams syndrome. *Developmental Cognitive Neuroscience* **29**:54-60 <https://doi.org/10.1016/j.dcn.2017.01.006> | PubMed
67. Liu H, Forest TA, Duncan K, Finn AS (2023) What sticks after statistical learning: The persistence of implicit versus explicit memory traces. *Cognition* **236** <https://doi.org/10.1016/j.cognition.2023.105439> | PubMed
68. Zwart FS, Vissers CThWM, Kessels RPC, Maes JHR (2018) Implicit learning seems to come naturally for children with autism, but not for children with specific language impairment: Evidence from behavioral and ERP data: Implicit learning intact in ASD but altered in SLI. *Autism Research* **11**:1050-61 <https://doi.org/10.1002/aur.1954> | PubMed
69. Kissine M (2021) Autism, constructionism, and nativism. *Language* **97**:e139-60 <https://doi.org/10.1353/lan.2021.0055>
70. Kuhl PK (2007) Is speech learning 'gated' by the social brain?. *Developmental Science* **10**:110-20 <https://doi.org/10.1111/j.1467-7687.2007.00572.x> | PubMed
71. Skoe E, Kraus N (2013) Musical training heightens auditory brainstem function during sensitive periods in development. *Front Psychol* **4** <https://doi.org/10.3389/fpsyg.2013.00622> | PubMed
72. Strait DL, Chan K, Ashley R, Kraus N (2012) Specialization among the specialized: Auditory brainstem function is tuned in to timbre. *Cortex* **48**:360-2 <https://doi.org/10.1016/j.cortex.2011.03.015> | PubMed
73. Yim D, Rudoy J (2013) Implicit statistical learning and language skills in bilingual children. *J Speech Lang Hear Res* **56**:310-22 [https://doi.org/10.1044/1092-4388\(2012/11-0243\)](https://doi.org/10.1044/1092-4388(2012/11-0243)) | PubMed
74. Karmiloff-Smith A (1998) Development itself is the key to understanding developmental disorders. *Trends in Cognitive Sciences* **2**:389-98 [https://doi.org/10.1016/S1364-6613\(98\)01230-3](https://doi.org/10.1016/S1364-6613(98)01230-3) | PubMed
75. Franchini M, Zöller D, Gentaz E, Glaser B, de Wilde H Wood, Kojovic N, et al. (2018) Early Adaptive Functioning Trajectories in Preschoolers With Autism Spectrum Disorders. *Journal of Pediatric Psychology* **43**:800-13 <https://doi.org/10.1093/jpepsy/jsy024> | PubMed
76. Nakuci J, Wasylyshyn N, Cieslak M, Elliott JC, Bansal K, Giesbrecht B, et al. (2023) Within-subject reproducibility varies in multi-modal, longitudinal brain networks. *Sci Rep* **13**:6699 <https://doi.org/10.1038/s41598-023-33441-3> | PubMed
77. Reuter M, Schmansky NJ, Rosas HD, Fischl B (2012) Within-subject template estimation for unbiased longitudinal image analysis. *NeuroImage* **61**:1402-18 <https://doi.org/10.1016/j.neuroimage.2012.02.084> | PubMed
78. Mullen EM (1995) *Mullen Scales of Early Learning* Circle Pines, MN: American Guidance Service.
79. Lord C, Risi S, DiLavore PS, Shulman C, Thurm A, Pickles A (2006) Autism From 2 to 9 Years of Age. *Arch Gen Psychiatry* **63**:694 <https://doi.org/10.1001/archpsyc.63.6.694> | PubMed
80. Shen MD, Nordahl CW, Young GS, Wootton-Gorges SL, Lee A, Liston SE, et al. (2013) Early brain enlargement and elevated extra-axial fluid in infants who develop autism spectrum disorder. *Brain* **136**:2825-35 <https://doi.org/10.1093/brain/awt166> | PubMed | PubMed Central
81. Ozonoff S, Young GS, Carter A, Messinger D, Yirmiya N, Zwaigenbaum L, et al. (2011) Recurrence risk for autism spectrum disorders: a Baby Siblings Research Consortium study. *Pediatrics* **128**:e488-495 <https://doi.org/10.1542/peds.2010-2825> | PubMed | PubMed Central
82. Sacrey LAR, Bryson S, Zwaigenbaum L, Brian J, Smith IM, Roberts W, et al. (2018) The Autism Parent Screen for Infants: Predicting risk of autism spectrum disorder based on parent-reported behavior observed at 6–24 months of age. *Autism* **22**:322-34 <https://doi.org/10.1177/1362361316675120> | PubMed
83. Van Nuland A, Reddy T, Quassem F, Vassalli JD, Berg AT (2021) PACS1-Neurodevelopmental disorder: clinical features and trial readiness. *Orphanet J Rare Dis* **16**:386 <https://doi.org/10.1186/s13023-021-02001-1> | PubMed

84. Agrawal S, Rao SC, Bulsara MK, Patole SK (2018) Prevalence of Autism Spectrum Disorder in Preterm Infants: A Meta-analysis. *Pediatrics* **142**:e20180134 <https://doi.org/10.1542/peds.2018-0134> | PubMed
85. Bayley N (1993) *BSID-II: Bayley Scales of infant development (Bsid-II)* San Antonio, TX: Psychological Corporation.
86. Dutoit T, Pagel V, Pierret N, Bataille F, Van Der Vrecken O (1996) The MBROLA project: towards a set of high quality speech synthesizers free of use for non commercial purposes. In: Proceeding of Fourth International Conference on Spoken Language Processing. ICSLP '96. Philadelphia, PA. pp. 1393-6 <https://doi.org/10.1109/ICSLP.1996.607874>
87. Fló A, Gennari G, Benjamin L, Dehaene-Lambertz G (2022) Automated Pipeline for Infants Continuous EEG (APICE): A flexible pipeline for developmental cognitive studies. *Developmental Cognitive Neuroscience* **54** <https://doi.org/10.1016/j.dcn.2022.101077> | PubMed
88. Delorme A, Makeig S (2004) EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods* **134**:9-21 <https://doi.org/10.1016/j.jneumeth.2003.10.009> | PubMed
89. Marriott Haresign I, Phillips E, Whitehorn M, Noreika V, Jones EJH, Leong V, et al. (2021) Automatic classification of ICA components from infant EEG using MARA. *Developmental Cognitive Neuroscience* **52** <https://doi.org/10.1016/j.dcn.2021.101024> | PubMed
90. Benjamin L, Fló A, Palu M, Naik S, Melloni L, Dehaene-Lambertz G (2021) Tracking transitional probabilities and segmenting auditory sequences are dissociable processes in adults and neonates. *bioRxiv* <https://doi.org/10.1101/2021.09.02.458702>
91. De Cheveigné A, Simon JZ (2008) Denoising based on spatial filtering. *Journal of Neuroscience Methods* **171**:331-9 <https://doi.org/10.1016/j.jneumeth.2008.03.015> | PubMed
92. McIntosh AR, Lobaugh NJ (2004) Partial least squares analysis of neuroimaging data: applications and advances. *NeuroImage* **23**:S250-63 <https://doi.org/10.1016/j.neuroimage.2004.07.020> | PubMed
93. Lobaugh NJ, West R, McIntosh AR (2001) Spatiotemporal analysis of experimental differences in event-related potential data with partial least squares. *Psychophysiology* **38**:517-30 <https://doi.org/10.1017/S0048577201991681> | PubMed
94. Delavari F, Rafi H, Sandini C, Murray RJ, Latrèche C, Van De Ville D, et al. (2023) Amygdala subdivisions exhibit aberrant whole-brain functional connectivity in relation to stress intolerance and psychotic symptoms in 22q11.2DS. *Transl Psychiatry* **13**:145 <https://doi.org/10.1038/s41398-023-02458-7> | PubMed
95. Kebets V, Holmes AJ, Orban C, Tang S, Li J, Sun N, et al. (2019) Somatosensory-Motor Dysconnectivity Spans Multiple Transdiagnostic Dimensions of Psychopathology. *Biological Psychiatry* **86**:779-91 <https://doi.org/10.1016/j.biopsych.2019.06.013> | PubMed
96. Zöllner D, Sandini C, Karahanoğlu FI, Padula MC, Schaer M, Eliez S, et al. (2019) Large-Scale Brain Network Dynamics Provide a Measure of Psychosis and Anxiety in 22q11.2 Deletion Syndrome. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging* **4**:881-92 <https://doi.org/10.1016/j.bpsc.2019.04.004> | PubMed
97. Geladi P, Kowalski BR (1986) Partial least-squares regression: a tutorial. *Analytica Chimica Acta* **185**:1-17 [https://doi.org/10.1016/0003-2670\(86\)80028-9](https://doi.org/10.1016/0003-2670(86)80028-9)
98. Zöllner D, Schaer M, Scariati E, Padula MC, Eliez S, Van De Ville D (2017) Disentangling resting-state BOLD variability and PCC functional connectivity in 22q11.2 deletion syndrome. *NeuroImage* **149**:85-97 <https://doi.org/10.1016/j.neuroimage.2017.01.064> | PubMed

Peer reviews

Reviewer #1 (Public review):

Summary:

This manuscript reports a prospective longitudinal study examining whether infants with high likelihood (HL) for autism differ from low-likelihood (LL) infants in two levels of word learning: brain-to-speech cortical entrainment and implicit word segmentation. The authors report reduced syllable tracking and post-learning word recognition in the HL group relative to the LL group. Importantly, both the syllable-tracking entrainment measure and the word recognition ERP measure are positively associated with verbal outcomes at 18-20 months, as indexed by the Mullen Verbal Developmental Quotient. Overall, I found this to be a thoughtfully designed and carefully executed study that tackles a difficult and important set of questions. With some clarifications and modest additional analyses or discussion on the points below, the manuscript has strong potential to make a substantial contribution to the literature on early language development and autism.

Strengths:

This is an important study that addresses a central question in developmental cognitive neuroscience: what mechanisms underlie variability in language learning, and what are the early neural correlates of these individual differences? While language development has a relatively well-defined sensitive period in typical development, the mechanisms of variability-particularly in the context of neurodevelopmental conditions-remain poorly understood, in part because longitudinal work in very young infants and toddlers is rare. The present study makes a valuable contribution by directly targeting this gap and by grounding the work in a strong theoretical tradition on statistical learning as a foundational mechanism for early language acquisition.

I especially appreciate the authors' meticulous approach to data quality and their clear, transparent description of the methods. The choice of partial least squares correlation (PLS-c) is well motivated, given the multidimensional nature of the data and collinearity among variables and the manuscript does a commendable job explaining this technique to readers who may be less familiar with it.

The results reveal interesting developmental changes in syllable tracking and word segmentation from birth to 2 years in both HL and LL infants. Simply mapping these trajectories in both groups is highly valuable. Moreover, the associations between neural indices of brain-to-speech entrainment and word segmentation with later verbal outcomes in the LL group support a critical role for speech perception and statistical learning in early language development, with clear implications for understanding autism. Overall, this is a rich dataset with substantial potential to inform theory.

Comment on revised version.

The revised manuscript has provided additional analyses that lead to critical clarification of the main findings, including the longitudinal nature of the relationship between neural tracking of speech and language, the role of sleep, and the potential modulation effect of stream structure on syllable-level neural tracking. The overall results highlight the robustness of the findings as well as the specific relevance of the structured speech tracking to verbal outcomes of infants with high likelihood (HL) of autism.

<https://doi.org/10.7554/eLife.109901.2.sa2>

Reviewer #2 (Public review):

Summary:

This article looks at differences in how the brain entrains to, or tracks, the rhythmic presentation of syllables and words in speech in infants at increased likelihood versus low likelihood for autism. The authors first sought to characterize how brain responses are

modulated by learning the statistical probability of a given syllable following the one before it over the first two years of life. They then sought to identify at which stages of word learning infants at increased likelihood for autism showed difficulties, and whether those difficulties worsened over time. Finally, they sought to indicate whether infants' statistical learning and word learning abilities could predict later verbal skills. The authors found similar developmental trajectories of neural entrainment to syllables in infants at high and low likelihood for autism, but infants at high likelihood for autism had overall weaker syllable-level entrainment. Infants at high versus low likelihood for autism showed different developmental trajectories for word entrainment. Lower syllable entrainment in high-likelihood infants corresponded with poorer verbal outcomes, but word entrainment was not associated with verbal outcomes. Event-related potential responses to words and part words were positively associated with verbal outcomes, however, but only in low-likelihood infants.

Strengths:

Overall, the article provides rigorous statistical analysis of longitudinal EEG data to provide strong support for the claims that neural entrainment to syllable and word features of speech may be a useful marker for language development difficulties, particularly in infants at increased likelihood for neurodevelopmental disorders. The EEG data collection and preprocessing procedures are well within standards within the field. Readers should take care to note that authors indexed neural entrainment to speech using phase-locking values instead of spectral power.

Comments on revised version.

While the statistical analyses are rigorous, there are a few potential confounds to the results. The authors now do a nice job addressing these limitations to the work. For example, sleep status may modulate some of the biomarkers relevant for language learning. Exposure to additional languages may influence performance on the verbal assessment, though the authors do clarify that participants came from majority French-speaking households. As a result, readers should be encouraged to interpret that neural entrainment to speech features is likely a useful mechanism to explain differences in language development, while taking this interpretation with some caution.

<https://doi.org/10.7554/eLife.109901.2.sa1>

Author response:

The following is the authors' response to the current reviews.

Reviewer #1 (Recommendations for the authors):

(1) Interpretation of Syllable-Tracking in the RND Condition:

The finding of greater syllable-tracking in the LL group compared to the HL group in the RND condition warrants cautious interpretation. Currently, there is no direct statistical evidence demonstrating greater PLV at 4 Hz in the Structured versus Random conditions for either group; readers must infer this solely from numeric differences in Figure S5 B and D. Therefore, while the interpretation on Page 14 (Lines 443-446) "successful segmentation may enhance syllable tracking via top-down predictions of the next syllable" is an interesting speculation, it feels somewhat far-reaching. Additionally, the authors should discuss whether this upregulated syllable tracking in the structured condition (which is specific to the HL group) represents an adaptive or maladaptive response.

The reviewer correctly highlights the lack of direct comparison between conditions (RND versus STR). We tempered our claims in the cited paragraph and insisted on the speculative

nature of this part of the discussion. We also clarified that, to us, it may represent an adaptive compensatory strategy:

Page 14, line 441: “Interestingly, our supplementary analyses (Supplementary Material Figure S4-5) suggest that syllable entrainment may be differentially affected in HL versus LL infants, depending on the statistical structure of the input stream (RND versus STR). However, as our experiment was not explicitly designed to test stream effects, these results should be interpreted with caution. Future studies could explore how successful segmentation may enhance syllable tracking via top-down predictions of the next syllable in both LL and HL infants. If confirmed, such a mechanism may improve alignment to syllable onsets, potentially constituting a compensatory process allowed by preserved segmentation abilities.”

(2) Preservation of Statistical Learning in HL Infants:

The text added on Pages 17-18 (Lines 562-566) regarding a "heightened dependence on bottom-up mechanisms (in autism)" does not appear to be supported by the data or by theories of implicit statistical learning. Because greater syllable-level entrainment was observed in the LL group than the HL group across both the random and structured conditions, the data actually point toward impaired bottom-up processes. Furthermore, implicit statistical learning typically involves an interplay of both bottom-up and top-down mechanisms; the implicit nature of a task does not guarantee a strictly bottom-up process. Consequently, this interpretation is not entirely convincing.

We agree with the reviewer that the concepts of “top-down” and “bottom-up” were not fully appropriate to support our point in the cited paragraph. We should have used the concepts of implicit versus explicit learning instead, in line with previous literature suggesting increased reliance on preserved implicit learning in autism to compensate for altered explicit processes. The paragraph was slightly modified.

Page 18, line 564: “According to these studies, autistic impairments in explicit attentional processes, such as social orienting - which are critical for bootstrapping language acquisition (70) - may result in a heightened dependence on implicit mechanisms, including statistical learning. As previously discussed, preserved word segmentation abilities may further compensate for alterations in lower-level implicit processes, such as syllable tracking.”

Reviewer #2 (Recommendations for the authors):

Potential typo on line 199 - I think an apostrophe is needed here.

Potential typo on line 255 - do you mean Central electrodes?

We addressed the typos spotted by reviewer.

Line 199: variables’

Line 255: Centro-frontal electrodes

The following is the authors’ response to the original reviews

Public Reviews:

Reviewer #1 (Public review):

Summary:

This manuscript reports a prospective longitudinal study examining whether infants with high likelihood (HL) for autism differ from low-likelihood (LL) infants in two levels of word learning: brain-to-speech cortical entrainment and implicit word segmentation. The

authors report reduced syllable tracking and post-learning word recognition in the HL group relative to the LL group. Importantly, both the syllable-tracking entrainment measure and the word recognition ERP measure are positively associated with verbal outcomes at 18-20 months, as indexed by the Mullen Verbal Developmental Quotient. Overall, I found this to be a thoughtfully designed and carefully executed study that tackles a difficult and important set of questions. With some clarifications and modest additional analyses or discussion on the points below, the manuscript has strong potential to make a substantial contribution to the literature on early language development and autism.

Strengths:

This is an important study that addresses a central question in developmental cognitive neuroscience: what mechanisms underlie variability in language learning, and what are the early neural correlates of these individual differences? While language development has a relatively well-defined sensitive period in typical development, the mechanisms of variability - particularly in the context of neurodevelopmental conditions - remain poorly understood, in part because longitudinal work in very young infants and toddlers is rare. The present study makes a valuable contribution by directly targeting this gap and by grounding the work in a strong theoretical tradition on statistical learning as a foundational mechanism for early language acquisition.

I especially appreciate the authors' meticulous approach to data quality and their clear, transparent description of the methods. The choice of partial least squares correlation (PLS-c) is well motivated, given the multidimensional nature of the data and collinearity among variables, and the manuscript does a commendable job explaining this technique to readers who may be less familiar with it.

The results reveal interesting developmental changes in syllable tracking and word segmentation from birth to 2 years in both HL and LL infants. Simply mapping these trajectories in both groups is highly valuable. Moreover, the associations between neural indices of brain-to-speech entrainment and word segmentation with later verbal outcomes in the LL group support a critical role for speech perception and statistical learning in early language development, with clear implications for understanding autism. Overall, this is a rich dataset with substantial potential to inform theory.

Weaknesses:

(1) Clarifying longitudinal vs. concurrent associations

Because the current analytical approach incorporates all time points, including the final visit, it is challenging to determine to what extent the brain-language associations are driven by longitudinal relationships vs. concurrent correlations at the last time point. This does not undermine the main findings, but clarifying this issue could significantly enhance the impact of the individual-differences results. If feasible, the authors might consider (a) showing that a model excluding the final visit still predicts verbal outcomes at the last visit in a similar way, or (b) more explicitly acknowledging in the discussion that the observed associations may be partly or largely driven by concurrent correlations. Either approach would help readers interpret the strength and nature of the longitudinal claims.

We thank the reviewer for this insightful comment. We agree that distinguishing between longitudinal predictive power and concurrent correlations at the final visit is crucial for clarifying the nature of these brain-language associations. Following the reviewer's suggestion (a), we re-ran the two critical Partial Least Squares Correlation (PLS-c) analyses by excluding all EEG and behavioral data from the final 18–21 month visit ($n = 54$ recordings kept) to test whether earlier trajectories still predict the final verbal outcome.

(1) Syllable entrainment (4 Hz) (original analysis on Figure 2C–D): The PLS-c restricted to the 3- to 15-month visits still identified a single significant component ($p = .001$, $r = .56$, 64.0% explained covariance, Figure 2-figure supplement 3). Bootstrap ratios (BSR) were: contrast (low vs. high autism likelihood) 4.1; mean age -1.3 ; contrast*mean-age -2.1 ; delta-age 10.5; contrast*delta-age 2.0; age² -6.0 ; contrast*age² -5.8 ; and notably verbal outcome 7.1; contrast*verbal-outcome -6.3 .

The latent component and its spatial electrode configuration remain highly consistent with the original analysis (Figure 2C–D). This confirms that excluding the final visit preserves the model's predictive validity: lower syllable entrainment correlates with poorer verbal outcomes at 18–21 months, particularly in the high-likelihood group.

(2) Late evoked response to novel words (original analysis on Figure 6): The PLS-c analysis on the ERP late time window (1500–3000 ms), excluding the final visit, also revealed one significant component ($p = .002$, $r = .74$, 33.5% explained covariance, Figure 6-figure supplement 1). Bootstrap ratios (BSR) were: contrast (part-word versus word) 14.5; mean age -5.2 ; contrast*mean-age 6.2; delta-age -0.8 ; contrast*delta-age 3.5; age² -0.8 ; contrast*age² -12.1 ; verbal-outcome 20.1; contrast*verbal-outcome -7.2 . The latent component closely mirrored the original analysis (Figure 6), with frontal electrodes contributing negatively and posterior electrodes positively. Minor divergences in age-related parameter contributions were observed, likely due to the absence of 18-21 month timepoints, which previously contributed to the convex/concave shapes of the group age trajectories in figure 6B (left panel).

Crucially, both models (with and without the final visit) positively predicted verbal outcomes (Figure 6B, right panel, and Figure 6-figure supplement 1B). However, excluding the final visit reversed the direction of the group*verbal-outcome interaction (from 3 to -7.2): This indicates that after ruling out cross-sectional correlations at 18–21 months, the early predictive value of the late ERP to word novelty is more prominently observed in high-likelihood infants, suggesting that the original result was influenced by concurrent cross-sectional correlations at the final visit. This aligns with the syllable entrainment findings (Figure 2-figure supplement 3), as both 4 Hz neural tracking and late ERP responses to novelty predominantly predict verbal outcomes in infants at high likelihood for autism.

We reported these supplementary analyses in the revised manuscript as follows:

We added Figure 2-figure supplement 3 and Figure 6-figure supplement 1. In general, most of figures that were present in Supplementary materials were moved as figure supplements to enhance readability.

Page 8 lines 234-240 (pages and lines refer to the reviewed uploaded manuscript): “To rule out the possibility that the association between syllable entrainment and verbal outcome was driven by concurrent measures taken at 18–21 months, we re-ran the PLS-c analysis excluding EEG data from the final visit ($n = 54$ recordings kept). The resulting latent component remain significant ($p = .001$) and showed contributions from behavioral and EEG variables that were highly similar to those observed in the previous analysis, with a verbal outcome BSR of 7.1 and a group*verbal-outcome interaction BSR of -6.3 (Figure 2-figure supplement 3).”

Page 12 lines 387-394: “As we did for neural entrainment to syllables, we conducted a new analysis on late ERP to word novelty, excluding EEG data from the final visit. This PLS-c yielded one significant latent component ($p = .002$, $r = .74$, 33.5% explained covariance, Figure 6 -figure supplement 1) with globally similar EEG parameter contributions and age trajectory modelling. Verbal outcome still significantly contributed to the latent component ($BSR = 20.1$), with a negative verbal outcome*group interaction ($BSR = -7.2$). These results suggest that, after ruling out cross-sectional correlations at 18–21 months, the late ERP to word novelty predominantly predicts verbal outcomes in high-likelihood infants for autism.”

Page 17 lines 547-548: “As with syllable entrainment, the late ERP to novel words primarily predicted verbal outcomes in high-likelihood (HL) infants.”

Page 18 lines 588-590: “Likewise, the absence of a late ERP orientation response in HL participants may represent an early neural signature of altered attention to novelty that can be used both as a non-invasive predictor of language development and as a potential target for early intervention.”

(2) Incorporating sleep status into longitudinal models

Sleep status changes systematically across developmental stages in this cohort. Given that some of the papers cited to justify the paradigm also note limitations in speech entrainment and word segmentation during sleep or in patients with impaired consciousness, it would be helpful to account for sleep more directly. Including sleep status as a factor or covariate in the longitudinal models, or at least elaborating more fully on its potential role and limitations, would further strengthen the conclusions and reassure readers that these effects are not primarily driven by differences in sleep-wake state.

The reviewer is highlighting here a limitation of our study design that comprised sleeping status that varied from one timepoint to another among participants. To rule out any confounding effect of wake status (coded as a binary variable: sleeping or awake during recording) on analyses comparing groups, a linear mixed-effect model with repeated measures was fitted finding no significant difference between high- and low-likelihood participants ($p = .769$, reported at page 20, lines 646-647). However, as rightly suggested by the reviewer, this doesn't prevent from a sleep bias on age trajectories, especially given that sleeping status significantly decreases with age in our sample.

Including sleep status as a covariate in our analyses, as suggested by the reviewer, would be difficult to implement in our PLS-c methods, since a categorical behavioral parameter that varies within participants is not possible in the models provided by myPLS toolbox.

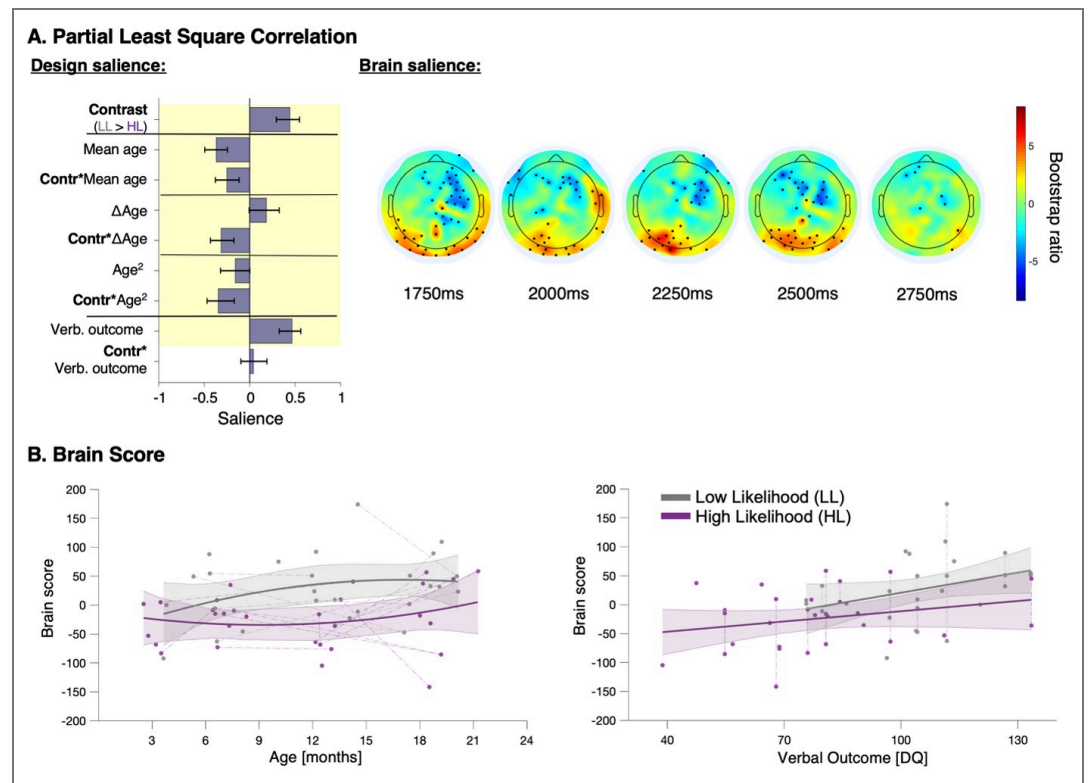
As an alternative option, we re-ran all analyses that explored the condition effect on the whole sample within the sleeping participants only ($n = 25$ recordings) to confirm that the same age-trajectories of EEG parameters were highlighted. However, negative results should be interpreted with caution since the sample is small for such a multivariate approach, resulting in modest statistical power.

(1) Syllable entrainment (4 Hz) (original analysis on Figure 2A–B): The PLS-c identified one significant component ($p < .001$, $r = .78$, 85.1% explained covariance, Figure 2 -figure supplement 2 and Figure 3 -figure supplement 1). Bootstrap ratios (BSR) were: contrast (4hz vs. adjacent frequencies) 30.3; mean age -2.9; contrast*mean-age -2.4; delta-age 3.8; contrast*delta-age 3.4; age² -1.1; contrast* -2.5. The spatial distribution of contributing electrodes globally matched that shown in Figure 2A. The high contrast BSR (30.3) confirms robust syllable entrainment in sleeping infants. Critically, the contrast*age² parameter contributed negatively to the latent component ($BSR = -2.5$), confirming that the convex age trajectory of syllabic entrainment (Figure 2B) is also present in the sleeping subsample.

(2) Word entrainment (1.3 Hz) (original analysis on Figure 3A–B): The PLS-c identified one significant component ($p < .001$, $r = .63$, 37.3% explained covariance, Author response image 1). Bootstrap ratios (BSR) were: contrast (1.3hz vs. adjacent frequencies) 24.9; mean age -7.5; contrast*mean-age -3.2; delta-age 2.9; contrast*delta-age 0.0; age² 1.5; contrast*age² 1.1. The spatial distribution of significant electrodes partially overlaps with the ones in the original analysis, primarily showing fronto-central positive contribution to the latent component. The high contrast BSR confirms a robust word entrainment in sleeping participants, in line with previous studies (e.g., Flò et al, Sci Rep, 2022). However, the lack of a significant contrast*age² suggests that the U-shape age trajectory illustrated on Figure 3 might be modulated by wakefulness or due to a lack of power in the present analysis. A non-significant trend towards a U-shape pattern with a 12-month nadir is visible in sleeping participants, but additional data from sleeping 18-21 months sleeping infants would be required to confirm or refute this trend.

(3) Early evoked response to novel words (original analysis on Figure 4): The PLS-c analysis on the ERP early time window (0–1000 ms) in sleeping participants revealed no significant component. The absence of early response to word novelty in sleeping participant might account for the lack of response observed in the whole sample, illustrated on Figure 4. To test this hypothesis, we conducted the same PLS-c in awake participants ($n=58$ recordings), which also yielded no significant latent component. This suggests that the lack of a measurable early response to word novelty observed in the whole sample is consistent across both sleeping and awake infants, and not driven by any of the two subsamples.

(4) Late evoked response to novel words (original analysis on Figure 5): The PLS-c analysis on the ERP late time window in sleeping participants revealed no significant component. This suggests that sleeping participants might present a reduced or even absent late response to novel words. Given this identified effect of sleep on late ERP response, we reran the PLS-c on the late ERP window using group as contrast (original analysis on figure 6), excluding the sleeping participants to avoid any confounds. This PLS-c revealed one significant component ($p = .006$, $r = .68$, 30.5% explained covariance, Author response image 1). Bootstrap ratios (BSR) were: contrast (low versus high likelihood) 7.6; mean age -5.9; contrast*mean-age -3.8; delta-age 2.5; contrast*delta-age -3.5; age² -2.5; contrast*age² -4.9; verbal-outcome 8.5; contrast*verbal-outcome -1.0. Behavioral parameters contribute to this latent component with similar magnitude and polarity as in the original analysis. Electrode contributions are also highly consistent, with frontal negative and posterior positive contributions. This confirms that sleeping participants, despite their potentially reduced late response, did not significantly bias the results presented in Figure 6.



Author response image 1. Late evoked response potential (ERP) to word novelty in awake participants. A. Design and brain saliences derived from the significant latent component. Brain topographies of bootstrap ratios (BSR) are displayed at 250ms intervals. Black dots indicate BSR > 2.3. **B.** Participants' brain scores for part-word and word conditions, as a function of age (left panel) and verbal DQ (right panel). For details on brain scores, see Figure 6 -figure supplement 1. Linear fitting is used for illustrative purposes only. HL: high likelihood for autism; LL: low likelihood for autism.

We reported these analyses in the revised manuscript as follows:

We added Figure 2 -figure supplement 2A and Figure 3 -figure supplement 1.

Page 7, lines 218-224: “Because some infants were asleep during the recording session, particularly at younger ages, we performed a supplementary control analysis restricted to this sleeping subsample (n = 25 recordings, Figure 2 -figure supplement 2). This PLS-c also identified a significant latent component (p < .001, r = .78, 85.1% explained covariance), with a significant contrast effect (BSR = 30.3) and a significant negative contrast*age² interaction (BSR = -2.5). These findings confirm that the convex age trajectory observed in the main analysis remains present and observable even in sleeping infants.”

Page 8, lines 252-259: “We further investigated word entrainment in sleeping participants (n=25), which yielded one significant latent component (p<.001, r=.63, 37.3% explained covariance, Figure 3 -figure supplement 1). Centro-frontal electrode contributed to this component, with a high contrast BSR (24.9), confirming a similar word entrainment pattern in the sleeping subsample. The contrast*age² was also positive but not significant (1.1), suggesting a trend toward a U-shape age trajectory with a 12-month nadir in sleeping infants. Additional 18-21 month recording would be required to confirm this trend.”

Page 11, lines 347-349: “The same PLS-c, conducted separately in sleeping (n=25) and awake subsamples (n = 58), yielded no significant latent component, indicating a consistent absence of early response to word novelty in both sleeping and awake infants.”

Page 11-12 lines 368-370: “The same PLS-c in the sleeping subsample yielded no significant latent component, suggesting that sleep may reduce or even abolish the late response to word novelty.”

Page 12 lines 382-385: “Given that no late response was detected in sleeping participants, we re-ran the PLS-c analysis using group as a contrast in the awake subsample (n=58). This yielded one significant latent component ($p=.006$, $r=.68$, 30.5% explained covariance), with behavioral and electrode contributions highly overlapping with those in Figure 6.”

Page 18 lines 590-592: “This potential biomarker might nevertheless be modulated by participants’ sleep status, warranting careful consideration of vigilance state in future studies.”

(3) Use of PLS-c and potential group $\times$ condition interactions

I am relatively new to PLS-c. One question that arose is whether PLS-c could be extended to handle a two-way interaction between group and condition contrasts (STR vs. RND). If so, some of the more complex supplementary models testing developmental trajectories within each group (Page 8, Lines 258-265) might be more directly captured within a single, unified framework. Even a brief comment in the methods or discussion about the feasibility (or limitations) of modeling such interactions within PLS-c would be informative for readers and could streamline the analytic narrative.

The reviewer raises a valid concern regarding the capacity of PLS-c to accommodate multi-way interactions among categorical and continuous variables. While PLS-c has no inherent theoretical constraints on the number of predictor terms (they can even exceed the sample size in number), practical limitations arise from model stability and interpretability when the ratio of predictors to sample size becomes excessive. As noted by Geladi and Kowalski (1986), exceeding ~10% of the sample size with predictors increases noise sensitivity and overfitting.

In our study, the PLS-c analyses already reach this ~10% limit, with a maximum of nine predictors for a sample size of $n=83$. Attempting to integrate both group and condition as contrasts — along with necessary age parameters to account for developmental trajectories — would result in 12 predictors (or 15 if verbal outcome is included). Specifically, the model would require behavioral terms for Group, Condition, Group*Condition, Mean-age, Group*Mean-age, Condition*Mean-age, Delta-age, Group*Delta-age, Condition*Delta-age, Age², Group*Age², Condition*Age², Verbal-outcome, Group*Verbal-outcome, and Condition*Verbal-outcome.

Although a unified multivariate model capturing the complex dynamics at play in our sample is theoretically appealing, the substantial risk of overfitting precludes its feasibility. Therefore, we opted to use only one categorical predictor per PLS-c analysis to maintain model parsimony and reliability. However, a larger sample could overcome this limitation, allowing a stable and unified model of longitudinal EEG data that simultaneously captures age trajectories, group, clinical outcome, and condition.

Reference:

Geladi, P., & Kowalski, B. (1986). Partial least-squares regression: A tutorial. *Analytica Chimica Acta*, 185, 1–17. [https://doi.org/10.1016/S0003-2670\(00\)82582-3](https://doi.org/10.1016/S0003-2670(00)82582-3)

We added the following comment in the method section:

Page 24, lines 773-777: “We limited the number of behavioral variables to nine to mitigate noise sensitivity and overfitting risks associated with exceeding the 10% sample size threshold (Geladi & Kowalski, 1986). This limitation precluded the implementation of a single PLS-c model incorporating group, condition (STR vs. RND), age, and their interactions.”

(4) STR-only analyses and the role of RND

Page 8, Lines 241-245: This analysis is conducted only within the STR condition. The lack of group difference observed here appears consistent with the lack of group difference in word-level entrainment (Page 9, Lines 292-294), suggesting that HL and LL groups may not differ in statistical learning per se, but rather in syllabic-level entrainment. As a useful sanity check and potential extension, it might be informative to explore whether syllable-level entrainment in the RND condition differs between groups to a similar extent as in Figure 2C-D. In other work (e.g., adults vs. children; Moreau et al., 2022), group differences can be more pronounced for syllable-level than for word-level entrainment. Figure S6 seems to hint that a similar pattern may exist here. If feasible, including or briefly reporting such an analysis could help clarify the asymmetry between the two learning measures and further support the interpretation of syllabic-level differences.

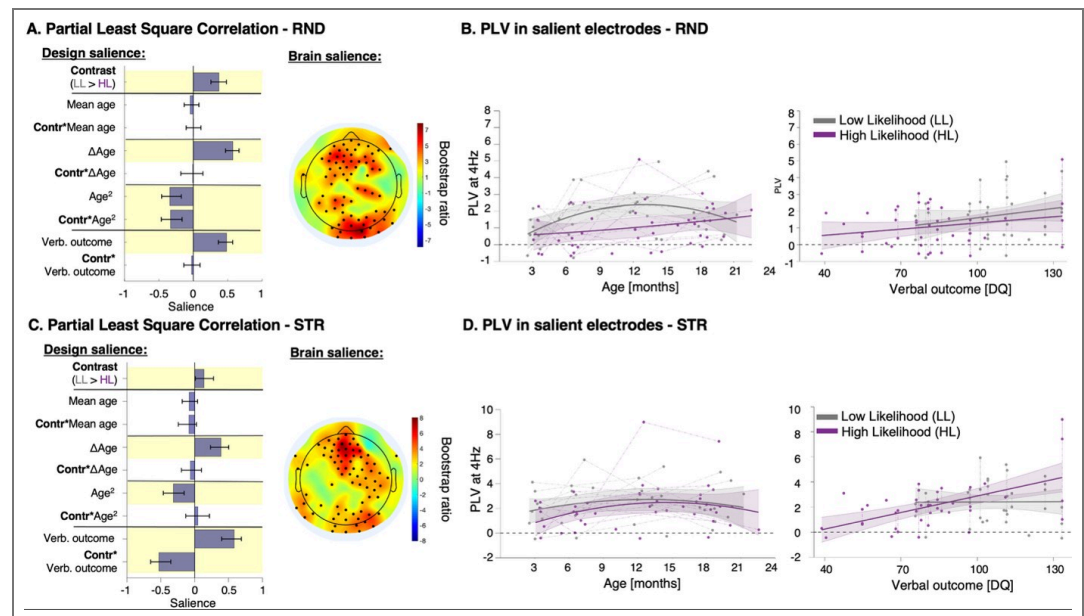
The reviewer points to the interesting pattern highlighted in supplementary figure S6, suggesting that group differences in syllabic entrainment might be modulated by the structure of the stream (STR versus RND). Such modulatory effect of stream structure on entrainment to syllables has been suggested by many studies, like Moreau et al (2022), as pointed by the reviewer, and seems at play in our sample, as illustrated on supplementary figure S5 (decline in the 4hz PLV that exceeds the size of confidence intervals, ~90 s after STR onset).

Following the reviewer’s suggestion, we ran a PLS-c testing group effect on 4hz PLVs in each stream:

(1) in the RND stream: the analysis yields one significant component ($p < .001$, $r = .49$, 52.7% explained covariance, Author response image 2A-B). Bootstrap ratios (BSR) are: contrast (low versus high likelihood) 6.9; mean age -1.0; contrast*mean-age 0.1; delta-age 11.8; contrast*delta-age 0.1; age² -4.9; contrast*age² 4.6; verbal-outcome 9.7; contrast*verbal-outcome -0.6. Interestingly, the model still highlights a strong link between syllable tracking and group, suggesting that RND also discriminate between HL and LL. However, RND syllable tracking doesn’t appear to be linked to group x verbal-outcome as we observed in Figure 2C-D.

(2) In the STR stream, we obtained one significant latent component ($p = .002$, $r = .51$, 57.8% explained covariance, Author response image 2C-D). Bootstrap ratios (BSR) are: contrast 2.2; mean age -1.6; contrast*mean-age -1.1; delta-age 6.1; contrast*delta-age -0.7; age² -4.4; contrast*age² 0.5; verbal-outcome 8.2; contrast*verbal-outcome -7.3. Here, the strong association between syllable tracking and group x verbal-outcome is similar to the model presented in Figure 2C-D.

Taken together, these results suggest that the apparent STR/RND dissociation illustrated in Figure S6 might primarily reflect a Group*Verbal-outcome divergence, with syllable tracking in the STR stream being related to verbal outcome mainly in high likelihood for autism.



Author response image 2. Syllable entrainment within RND (A-B) and STR (C-D).

These results were reported in the revised manuscript in the Result section (Time course of the entrainment along experiment subheader), implying a slight reframing of the result presentation of supplementary analysis S6. Author response image 2 was added in supplementary material as Figure S5.

Page 10, lines 307-319: “The group, age and verbal outcome parameters were mainly correlated (BSR>2.3) with the neural entrainment occurring~90 seconds after the onset of the STR stream, coinciding with the time participants began tracking word boundaries (Supplementary material, S3). This result suggests that the group differences in syllable entrainment, as shown in Figure 2C-D, as their associations with verbal outcome, are modulated by the structure of the stream (STR versus RND). We ran one additional PLS-c for each stream separately, using group as contrast. In both streams, the PLS-c yielded a significant LC ($p<.001$ for RND and $p=.002$ for STR), with a positive group effect (BSR>2.3) in both LC (Supplementary material, S5). Most strikingly, the group*verbal outcome parameter reached significance exclusively within the STR latent component (BSR:-7.3). These results suggest that while syllable tracking is generally decreased in HL infants across both streams, its association with verbal outcome is prominently driven by the stream containing words (STR).”

Page 14, lines 443-446: “This temporal overlap suggests that successful segmentation may enhance syllable tracking via top-down predictions of the next syllable, improving alignment to syllable onsets in LL infants as well as in HL with better verbal outcome.”

(5) Multi-speaker input and voice perception (Page 15, Lines 475-483)

The multi-speaker nature of the speech input is an interesting and ecologically relevant feature of the design, but it does add interpretive complexity. The literature on voice perception in autism is still mixed: for example, Boucher et al. (2000) reported no differences in voice recognition and discrimination between children with autism and language-matched non-autistic peers, whereas behavioral work in autistic adults suggests atypical voice perception (e.g., Schelinski et al., 2016; Lin et al., 2015). I found the current interpretation in this paragraph somewhat difficult to follow, partly because the data do not directly test how HL and LL infants integrate or suppress voice

information. I think the authors could strengthen this section by slightly softening and clarifying the claims.

We acknowledge the reviewer's concern regarding the potential ambiguity in the cited paragraph. To address this, we have revised the text to explicitly clarify the aims of our study and its design. Furthermore, we now emphasize the speculative and post-hoc nature of the hypotheses and interpretations presented, thereby ensuring transparency regarding the limitations of our findings.

Page 16 lines 520-530), as follows: "HL infants, on the other hand, did not show this transient disruption. In this group, word entrainment remained stable over time. To account for this unexpected finding, we followed up on the post-hoc hypothesis proposed above: a reduced sensitivity to social and vocal cues observed in HL infants may have spared segmentation abilities by limiting the interference introduced by speaker variability. If this post-hoc hypothesis holds true, LL and HL infants would differ not in their intrinsic ability to learn statistical regularities per se, but rather in how they integrate or suppress competing cues (such as speaker changes) during the segmentation process. It is important to note, however, that the present study was not designed to isolate and evaluate the specific impact of speaker changes on word segmentation. Consequently, this interpretation remains speculative, and additional research is required to further address this question."

(6) Asymmetry between EEG learning measures

Page 16, Lines 502-507 touches on the asymmetry between the two EEG learning measures but leaves some questions for the reader. The presence of word recognition ERPs in the LL group suggests that a failure to suppress voice information during learning did not prevent successful word learning. At the same time, there is an interesting complementary pattern in the HL group, who show LL-like word-level entrainment but does not exhibit robust word recognition. Explicitly discussing this asymmetry - why HL infants might show relatively preserved word-level entrainment yet reduced word recognition ERPs, whereas LL infants show both - would enrich the theoretical contribution of the manuscript.

We concur with the reviewer's observation that our findings imply a theoretically significant double dissociation between HL and LL groups, specifically concerning the asymmetries between word-level neural entrainment and word recognition mechanisms. We believe this point was partly addressed in the subsequent paragraph, where we stated that "in contrast" to LL, HL infants "showed no clear ERP difference between novel and familiar triplets", while "both groups showed similar word neural entrainment during learning". We further explored potential explanations for this apparent dissociation, such as a possible deficit in novelty orientation that may be specific to HL infants and unrelated to statistical learning itself. We cited Liu et al (2023) as a reference showing the dissociation between mechanisms underlying implicit versus explicit traces of statistical learning. We acknowledge that we can discuss more in depth the potential preservation of statistical learning in HL infants. We have incorporated the following discussion in the reviewed manuscript, supported by relevant references:

Pages 17-18, lines 562-566: "Interestingly, this dissociation between spared implicit versus impaired explicit statistical learning in autism has been previously discussed in the literature (Zwart et al, 2018, Kissine, 2021). According to these studies, autistic impairments in top-down attentional processes, such as social orienting — which are critical for bootstrapping language acquisition (Kuhl, 2007) — may result in a heightened dependence on bottom-up mechanisms, including implicit statistical learning."

References:

Zwart, F.S., Vissers, C.T.W.M., Kessels, R.P.C. and Maes, J.H.R. (2018), Implicit learning seems to come naturally for children with autism, but not for children with specific language impairment: Evidence from behavioral and ERP data. *Autism Research*, 11: 1050-1061. <https://doi.org/10.1002/aur.1954>

Kissine, M. (2021). Autism, constructionism, and nativism. *Language* 97(3), e139-e160. <https://dx.doi.org/10.1353/lan.2021.0055>.

Kuhl, P.K. (2007), Is speech learning 'gated' by the social brain?. *Developmental Science*, 10: 110-120. <https://doi.org/10.1111/j.1467-7687.2007.00572.x>

References:

(1) Moreau, C. N., Joanisse, M. F., Mulgrew, J., & Batterink, L. J. (2022). No statistical learning advantage in children over adults: Evidence from behaviour and neural entrainment. *Developmental Cognitive Neuroscience*, 57, 101154. <https://doi.org/10.1016/j.dcn.2022.101154>

(2) Boucher, J., Lewis, V., & Collis, G. M. (2000). Voice processing abilities in children with autism, children with specific language impairments, and young typically developing children. *Journal of Child Psychology and Psychiatry*, 41(7), 847-857. <https://doi.org/10.1111/1469-7610.00672>

(3) Schelinski, S., Borowiak, K., & von Kriegstein, K. (2016). Temporal voice areas exist in autism spectrum disorder but are dysfunctional for voice identity recognition. *Social Cognitive and Affective Neuroscience*, 11(11), 1812-1822. <https://doi.org/10.1093/scan/nsw089>

(4) Lin, I.-F., Yamada, T., Komine, Y., Kato, N., Kato, M., & Kashino, M. (2015). Vocal identity recognition in autism spectrum disorder. *PLOS ONE*, 10(6), e0129451. <https://doi.org/10.1371/journal.pone.0129451>

Reviewer #2 (Public review):

Summary:

This article looks at differences in how the brain entrains to, or tracks, the rhythmic presentation of syllables and words in speech in infants at increased likelihood versus low likelihood for autism. The authors first sought to characterize how brain responses are modulated by learning the statistical probability of a given syllable following the one before it over the first two years of life. They then sought to identify at which stages of word learning infants with increased likelihood of autism showed difficulties, and whether those difficulties worsened over time. Finally, they sought to indicate whether infants' statistical learning and word learning abilities could predict later verbal skills. The authors found similar developmental trajectories of neural entrainment to syllables in infants at high and low likelihood for autism, but infants at high likelihood for autism had overall weaker syllable-level entrainment. Infants at high versus low likelihood for autism showed different developmental trajectories for word entrainment. Lower syllable entrainment in high-likelihood infants corresponded with poorer verbal outcomes, but word entrainment was not associated with verbal outcomes. Event-related potential responses to words and part words were positively associated with verbal outcomes, however, but only in low-likelihood infants.

Strengths:

Overall, the article provides rigorous statistical analysis of longitudinal EEG data to provide strong support for the claims that neural entrainment to syllable and word

features of speech may be a useful marker for language development difficulties, particularly in infants at increased likelihood for neurodevelopmental disorders. The EEG data collection and preprocessing procedures are well within standards in the field. Readers should take care to note that authors indexed neural entrainment to speech using phase-locking values instead of spectral power.

Weaknesses:

While the statistical analyses are rigorous, a few of the components of the models are not clearly defined, and some corrections and thresholds for significance warrant further justification. Further, a few stimuli and participant details that could influence results are not specified. It is not clear whether all participants came from majority French-speaking families; differences in the amount of French language exposure (compared to other languages that may be spoken by a participant's family) could influence results. The standardized volume of the stimuli is also not included. As a result, readers should be encouraged to interpret that neural entrainment to speech features is likely a useful mechanism to explain differences in language development, while taking this interpretation with some caution.

We thank the reviewer for these remarks.

Regarding the amount of French exposure: while all participants were raised in primarily French-speaking environments (i.e., French as the dominant language at home and daycare), the parental questionnaire at intake indicated that 45% of the sample was exposed to additional languages, reflecting Geneva's highly multicultural demographics. We did not quantify the extent of this exposure, which could range from very occasional exposure to situations close to true bilingualism. The structural sensitivity hypothesis (Weiss et al., 2020) posits that additional language exposure may enhance detection of statistical structures in artificial language input, even when these structures differ from those in native languages. Yet, empirical support is mixed: Yim & Rudoy (2013) found no bilingualism effect in a paradigm close to ours (triplet segmentation via auditory statistical learning, n=112 children), whereas most studies reporting bilingual advantages for statistical learning involved tasks very distinct from ours, like artificial grammar and phonotactic rule learning, or multi-cue integration for segmentation (Weiss et al., 2020).

Regarding the volume of stimuli, they were played at 50cm distance with an intensity of 75dB. Both considerations have been included in the new version of the manuscript. In general, we moved most of the figures present in Supplementary material to figure supplements to improve readability.

Recommendations for the authors:

Reviewer #1 (Recommendations for the authors):

Minor Comments:

Figure 6: The figure caption is not complete (there is no description for the right half of panel B).

We thank the reviewer for this observation, figure 6 caption has been completed.

Reviewer #2 (Recommendations for the authors):

Broadly speaking, I would recommend reducing the number of abbreviations in this article, and I would recommend that the authors take care as to where these abbreviations are being introduced. Many of the abbreviated terms are defined in the Materials and Methods section, which is presented after the abbreviations are used in the main results.

We acknowledge that our extensive use of abbreviations compromises the readability of the manuscript. Consequently, we have removed the following abbreviations:

- SL (replaced by statistical learning)
- LC (replaced by latent component)
- ASD (replaced by autism)
- TP (replaced by transition probability)
- MEG (replaced by magneto-encephalogram)
- MSEL (replaced by Mullen Scale of Early Learnings)

The remaining abbreviations are:

HL (high likelihood for autism), LL (low likelihood for autism), EEG (electroencephalogram), PLS-c (partial least square correlation), ERP (event-related potential), RND (random), STR (structured), BSR (bootstrap ratio), AIC (Akaike Information Criterion), PLV (phase locking value), DQ (developmental quotient), APSI (Autism Parent Screen for Infants).

Moreover, we carefully reviewed how abbreviations were introduced and identified that PLS-c, STR and RND were not defined prior to the Method section. This oversight has been corrected in the reviewed manuscript.

I would also recommend that the authors be careful with the structuring of the Introduction, particularly with their research questions and hypotheses. The article initially makes clear that the research questions are focused on the developmental trajectory of statistical learning, the levels of word learning that may differentiate high-likelihood versus low-likelihood infants, and the stability of those differences, and associations between statistical learning and various levels of word learning with verbal outcomes. The use of acoustic variability across syllables, while a valuable methodological tool, is somewhat presented as an additional research question, but not clearly stated or tested as such.

We acknowledge that the introduction (particularly the paragraph from lines 173 to 184) may have implied that speaker variability across syllables was one of our primary research aims. We clarify here that speaker variability was introduced as a mean to increase task difficulty, particularly for high-likelihood (HL) participants, with the aim of amplifying the effect sizes in our analyses.

To address this, we have removed the theoretical discussion on speaker variability in autism and typical development (lines 173–184) and explicitly stated that speaker variability was not a research question in this study. Crucially, our experimental design did not include a control condition without speaker variability, and thus we could not test its specific effects on statistical learning across age trajectories and groups.

Page 6, lines 173-176 (pages and lines refer to the reviewed uploaded manuscript): “It is worth noting, however, that our study was not designed to isolate or quantify the specific impact of speaker variability on statistical learning, as the experimental design did not include a baseline control condition omitting this acoustic variation.”

The authors do a nice job in the Materials & Methods explaining PLS-c and defining the latent components and bootstrapped ratios that will be shared in the Results. An additional brief iteration defining these statistical elements is needed at the beginning of the Results section.

We thank the reviewer for their appreciation of our Method section. We agree that an additional iteration in the result section would improve readability. We added the following paragraph at the very beginning of the Result section, briefly defining PLS-c and its main statistical output (latent components and bootstrap ratios):

Pages 6-7, lines 193-202: “Briefly, PLS-c is a data-driven multivariate modelling approach designed to identify significant patterns of electrode clusters (from a brain data matrix containing electrophysiological measures, here PLV) and their associations with “behavioral” variables (from a behavioral design matrix, here age-related parameters). Patterns of brain x behavior associations are called latent components, and their statistical significance is evaluated using permutation testing ($n=1000$, Bonferroni correction for number of components tested, $\alpha=.006$). Brain and behavioral variables respective contributions to any significant latent component are tested with bootstrapping (500 random samples and replacement), with bootstrap ratios (BSR) greater than 2.3 indicating a stable contribution (for details, see the Materials and Methods section).”

(1) Page 18 Line 576. The authors need to clarify whether participants were required to be in primarily French-speaking environments and whether there was a minimum amount of French language exposure that participants were required to have if they were exposed to additional languages besides French in their everyday life.

The reviewer raises a valid concern regarding participants’ language exposure. In this study, all participants were raised in primarily French-speaking environments, with French as the dominant language at home and daycare. The parental questionnaire at intake indicated that 45% of the sample was exposed to additional languages, reflecting Geneva’s highly multicultural demographics. However, we did not quantify the extent of this exposure, which could range from very occasional exposure to situations close to true bilingualism.

The structural sensitivity hypothesis (Weiss et al., 2020) posits that additional language exposure may enhance detection of statistical structures in artificial language input, even when these structures differ from those in native languages. Yet, empirical support is mixed: Yim & Rudoy (2013) found no bilingualism effect in a paradigm close to ours (triplet segmentation via auditory statistical learning, $n=112$ children), whereas most studies reporting bilingual advantages for statistical learning involved tasks very distinct from ours, like artificial grammar and phonotactic rule learning, or multi-cue integration for segmentation (Weiss et al., 2020).

To include these considerations, Limitations and Material and methods sections were modified as follows:

Page 19, lines 604-607: “Second, although all participants were primarily exposed to French, we did not quantify additional language exposure, precluding any analysis of its potential moderator effects on statistical learning in our groups and age-trajectories. However, prior work has reported no effect of bilingualism on auditory triplet segmentation in children (Yim & Rudoy, 2013).”

Page 20, lines 630-631: “All participants were raised in primarily French-speaking environments, with French as the dominant language at home and daycare.”

References:

Weiss DJ, Schwob N, Lebkuecher AL. Bilingualism and statistical learning: Lessons from studies using artificial languages. *Bilingualism: Language and Cognition*. 2020;23(1):92-97. doi:10.1017/S1366728919000579

Yim D, Rudoy J. Implicit statistical learning and language skills in bilingual children. *J Speech Lang Hear Res*. 2013 Feb;56(1):310-22. doi: 10.1044/1092-4388(2012/11-0243). Epub 2012 Aug 15. PMID: 22896046.

(2) Page 18 Line 588. Further, the authors should clarify whether the 7 infants in the HL group, due to early parental concerns were defined by the 18-21-month APSI scores or by parental report prior to study enrollment.

These 7 infants were recruited based on early parental concerns prior to intake. The APSI score at 18-21 months is only reported to provide an illustration of the amount of early autistic signs that were present in these 7 infants, and to provide an estimation of their probability to develop autism later on based on Sacrey et al., 2018 longitudinal study on the APSI predictive value. We agree with the reviewer that our phrasing suggests that the APSI was used as an inclusion criterion. We rephrased the page 20 lines 642-646 as follows:

“The 7 other HL infants presented with early parental concerns for autism, based on parental report prior to enrollment. Their Autism Parent Screen for Infants (APSI) total score at their 18-21 months visit was 15.6 ± 6.4 , [8-22] range – a score greater than 8 reflecting a 63% positive predictive value for autism in HL populations.”

(3) Page 20 Line 641. The authors should specify the volume of the stimuli.

The volume of stimuli was reported in the main text (page 22, lines 695-696) as follows:

“Stimuli were played on a Bose® Companion 2 Series III at a 50cm distance with an intensity of 75dB.”

(4) I'd prefer Figure 1 to be reorganized slightly - at present, the placement of the arrows explaining the analysis steps is not intuitive.

We addressed the reviewer’s comments (4) and (5) together as they both refer to Figure 1B.

(5) Page 23 lines 718-719. I think it would be helpful to explicitly define each of the interaction variables included in the behavioral design matrix. Further, this matrix should be labeled consistently in both Figure 1B and in the main text.

We refined figure 1B and its corresponding main text (in Methods section) for clarity. The arrows are now simpler and more parsimonious, labels (e.g., participant i , visit n , behavior design matrix and its parameters) are now standardized between the figure and the main text, and the interaction terms at lines 718-719 are explicitly defined.

(6) Page 23 lines 726-731: It would be helpful to know whether applying a Bonferroni correction in addition to completing permutation testing is standard when evaluating latent components derived from PLS-c. The authors should also cite justification for a bootstrap ratio cutoff of 2.3 for defining stability.

In PLS-c analyses, multiple comparisons correction across latent components and bootstrap ratio (BSR) thresholding at 2.3 are commonly adopted practices.

- Correction for multiple comparisons in PLS-c: PLS-c performs singular decomposition of the data into latent components equal in number to the variables included in the behavior design matrix (7-9 in our study, depending on the inclusion of Verbal outcome as an input variable). Each latent component's statistical significance is assessed through permutation testing, generating a null distribution for its singular value (Krishnan et al., 2011). Given the multiple tests (one permutation test per latent component), Type I error inflation must be addressed. Recent PLS-c studies commonly applied Bonferroni correction (default procedure in the myPLS toolbox, used by Zoeller et al., 2017, and Delavari et al, 2021), though FDR correction has also been used (Lombardo et al, 2018).

- Stability threshold for bootstrap and replacement: Within each latent component, saliences' stability (brain/behavior parameter contributions to each latent component) are evaluated using bootstrapping (Krishnan et al., 2011). The bootstrap ratio (BSR) of each parameter, calculated as the saliency divided by its bootstrap-derived standard error, functions analogously to a z-score under normality assumptions. The BSR can then be used to assess the stability of the saliency (i.e., how stable is its contribution to the latent component). BSR thresholds in the literature typically range from 1.96 to 3.0. Krishnan et al (2011) state that when BSR are "larger than 2 the corresponding saliences are considered significantly stable". Delavari et al (2021) and our study used a 2.3 thresholding, corresponding to a 99.0% bootstrap confidence interval not crossing the zero line – roughly equivalent to a two-tailed $p < .001$. Lombardo et al (2018) used a looser threshold of 1.96, corresponding to a 95% confidence interval not crossing the zero line (~two-tailed $p < .05$), while Zöllner et al (2017) used a more stringent 3.0 thresholding (~ $p < .001$, or 99.9% confidence interval not crossing the zero line).

Thus, our application of Bonferroni correction for multiple comparisons and our 2.3 BSR threshold aligns with established conventions.

We added following lines in the manuscript:

Page 25 lines 784-785: "Bonferroni correction was applied to account for multiple comparisons across the 9 tested latent components in the PLS-c, yielding an adjusted alpha of .006 (Zoeller et al, 2017; Delavari et al, 2021)."

Page 25 lines 789-792: "BSR are analogous to Z-scores and can be used to assess the stability of the saliency. We considered $BSR > 2.3$ as stable, corresponding to a 99.0% bootstrap confidence interval not crossing zero – roughly equivalent to a two-tailed $p < .001$ (Delavari et al., 2021; Krishnan et al., 2011)."

References:

Delavari F, Sandini C, Zöllner D, Mancini V, Bortolin K, Schneider M, Van De Ville D, Eliez S. Dysmaturation Observed as Altered Hippocampal Functional Connectivity at Rest Is Associated With the Emergence of Positive Psychotic Symptoms in Patients With 22q11 Deletion Syndrome. *Biol Psychiatry*. 2021 Jul 1;90(1):58-68. doi: 10.1016/j.biopsych.2020.12.033. Epub 2021 Jan 18. PMID: 33771350.

Lombardo, M.V., Pramparo, T., Gazestani, V. et al. Large-scale associations between the leukocyte transcriptome and BOLD responses to speech differ in autism early language outcome subtypes. *Nat Neurosci* 21, 1680–1688 (2018). <https://doi.org/10.1038/s41593-018-0281-3>

Daniela Zöllner, Marie Schaer, Elisa Scariati, Maria Carmela Padula, Stephan Eliez, Dimitri Van De Ville. Disentangling resting-state BOLD variability and PCC functional connectivity in 22q11.2 deletion syndrome. *NeuroImage*, Volume 149, 2017, Pages 85-97, ISSN 1053-8119, <https://doi.org/10.1016/j.neuroimage.2017.01.064>

Anjali Krishnan, Lynne J. Williams, Anthony Randal McIntosh, Hervé Abdi, Partial Least Squares (PLS) methods for neuroimaging: A tutorial and review, *NeuroImage*, Volume 56, Issue 2, 2011, Pages 455-475, ISSN 1053-8119, <https://doi.org/10.1016/j.neuroimage.2010.07.034> 

| (7) I have a few minor grammar/formatting recommendations for the authors as well:

| (a) Should the Geneva Autism Cohort be capitalized? At present, it is not.

We agree with the reviewer's suggestion, and we capitalized the Geneva Autism Cohort in the main text (page 18, line 571)

| (b) Page 24, line 750. Do the authors mean that the data was re-referenced to average?

The preprocessed data is not average-referenced (see section Data pre-processing). Therefore, both for neural entrainment computation and ERPs, the data were average-referenced.

| (c) It would be nice to have a figure of the actual ERP for each condition and age group.

We agree that PLS-c can be difficult to interpret without the raw actual ERPs on which it was modelled. We direct the reviewer to supplementary figure S6 at page 59, which displays the raw ERPs for each condition (part-word, word, and their subtraction) per age group. Supplementary figures S7-8 at pages 60-61 further illustrate topographical ERPs for each group (high and low likelihood for autism). We deemed these figures too extensive for the main text. Instead, the most relevant ERP topographies are presented in Figures 4-6 to facilitate PLS-c interpretation.

<https://doi.org/10.7554/eLife.109901.2.sa0>