Attenuated processing of vowels in the left hemisphere predicts speech-in-noise perception deficit in children with autism.

Fadeev K.A.[1★], Romero Reyes, I.V.[1★], Goiaeva D.E.[1], Obukhova T.S.[1], Ovsiannikova T.M.[1], Prokofyev A.O.[1], Rytikova A.M.[1], Novikov A.Y.[2], Kozunov V.V.[1], Stroganova T.A.[1], Orekhova E.V.[1*]

[1] Center for Neurocognitive Research (MEG Center), Moscow State University of Psychology and Education, Moscow, Russian Federation
[2] RNSO "Center for Curative Pedagogics", Moscow, Russian Federation

★ The first two authors contributed equally to the work.
* Corresponding author

ABSTRACT

Background: Difficulties with speech-in-noise perception in autism spectrum disorders (ASD) may be associated with impaired analysis of speech sounds, such as vowels, which represent the fundamental phoneme constituents of human speech. Vowels elicit early (< 100 ms) sustained processing negativity (SPN) in the auditory cortex that reflects the detection of an acoustic pattern based on the presence of formant structure and/or periodic envelope information (f_0) and its transformation into an auditory 'object'.

Methods: We used magnetoencephalography (MEG) and individual brain models to investigate whether SPN is altered in children with ASD and whether this deficit is associated with impairment in their ability to perceive speech in the background of noise. MEG was recorded while boys with ASD and typically developing boys passively listened to sounds that differed in the presence/absence of f_0 periodicity and formant structure. Word-in-noise perception was assessed in the separate psychoacoustic experiment using stationary and amplitude modulated noise with varying signal-to-noise ratio.

Results: SPN was present in both groups with similarly early onset. In children with ASD, SPN associated with processing formant structure was reduced predominantly in the cortical areas lateral to and medial to the primary auditory cortex, starting at ~ 150 - 200 ms after the stimulus onset. In the left hemisphere, this deficit correlated with impaired ability of children with ASD to recognize words in amplitude-modulated noise, but not in stationary noise.

Limitations: Our analysis was limited to the auditory cortical areas, while ASD vs TD differences in vowel processing may exist beyond these areas. It is not clear if the SPN group differences are specific to vowels or are also present for other acoustic patterns with coherent frequency structure.

Conclusions: These results suggest that perceptual grouping of vowel formants into phonemes is impaired in children with ASD and that, in the left hemisphere, this deficit contributes to their difficulties with speech perception in fluctuating background noise.

Keywords: autism spectrum disorder (ASD), magnetoencephalography (MEG), speech-in-noise perception, vowels, formant structure, periodicity pitch, sustained processing negativity (SPN), children, auditory processing disorder

List of abbreviations

ASD - autism spectrum disorder
AM - amplitude-modulated [noise]

EEG - electroencephalography
ERP/ERF - event-related potential/field
INS - insula
MEG - magnetoencephalography
MMN/MMF - mismatch negativity/field
MPI - Mental Processing Index
ROI - region of interest
pINS - posterior insula
bcROI - ROI based on between-conditions difference
bgROI - ROI based on groups × conditions difference
SCQ - Social Communication Questionnaire
sLORETA - standardized low-resolution brain electromagnetic tomography
SNR - signal-to-noise ratio
SPN - sustained processing negativity
SPNadj - SPN adjusted for the response to the test stimulus and for the number of averaged epochs.
SRS - Social Responsiveness Scale
ST - stationary [noise]
STG - superior temporal gyrus
STS - superior temporal sulcus
TD - typically developing
TFCE - Threshold-Free Cluster Enhancement
WiN - words in noise
WiNam - words in amplitude-modulated noise
WiNst - words in stationary noise

BACKGROUND

Autism spectrum disorder (ASD) is a group of neurodevelopmental conditions characterized by social and communication impairments and repetitive and restricted behaviors and interests [1]. Most individuals with ASD have delayed early language development, and deficits in receptive and expressive language present to some degree in the majority of children with ASD [2,3]. A common difficulty faced by people with ASD, even those with normal or above normal IQ, is a poor listening ability under suboptimal acoustic conditions, both in experimental settings (for a review see [4]) and in real-life [5].

It is still a matter of debate whether receptive language deficits in people with ASD relates to the level of basic auditory speech processing or, especially in high-functioning individuals, are present only in the domains of pragmatics, syntax, and semantics [6]. Another unresolved question is to what extent basic speech processing abnormalities, if exist, could be secondary to disturbed cognitive functions such as attention and executive control or insufficient language experience (for a review see [7,8]).

Studies of vowel processing may shed light on some of these issues. Vowels represent the simplest and ontogenetically and phylogenetically oldest phoneme constituents of human speech. They are the first speech sounds to be produced by human infants [9]. Vowel-like sounds are present in the vocal repertoire of non-human primates [10]. Processing of vowel sound in the auditory pathways leads to the extraction of complex features (formant structure, common periodicity, etc.) from an acoustic pattern to further process the vowel as an auditory 'object' [11–13]. Detection of acoustic patterns is a rapid and automatic process that, in humans and other primates, is subserved by the auditory cortex [14–17].

The combination of formants, i.e., peaks in the frequency spectrum, determines the identity of the vowel, and the periodicity of the amplitude envelope (f0 / pitch) is important for speaker identification [18]. In the absence of linguistic context (i.e., when the vowel is not represented as part of a word), the spectral-temporal structure of the vowel remains the only auditory cue for the bottom-up grouping that governs its neural representation as a perceptually meaningful auditory object. There is some evidence that the ability to automatically group sound features is reduced in ASD [19–21]. It has also been suggested that this deficit may contribute to impaired speech perception in noisy environments in individuals with ASD, as the auditory system must rely on automatic grouping to effectively process speech [20]. However, whether automatic vowel processing in the auditory cortex in autistic individuals is altered in a way that affects speech perception remains an open question.

The processing of auditory patterns leading to formation of auditory objects (figures) is associated with a sustained negative shift in neural current recorded with MEG/EEG [16,17,22–25], which we will refer to here as sustained processing negativity (SPN). It has been suggested that SPN reflects the persistent activity of 'non-synchronized' neural populations [25,26]. These neurons are most abundant in non-primary auditory areas, are highly sensitive to complex sound features such as combinations of certain spectral and temporal parameters, and are thought to support representation of meaningful auditory patterns including species-specific vocalizations or other ecologically relevant sounds [13–15,27–29]. The term 'non-synchronized' refers to the property of these neurons to sustained firing for hundreds of milliseconds, especially when driven by their preferred continuous stimuli. These neurons are thought to encode temporally integrated spectral information [14,15,30] and transform it from an acoustic to a perceptual dimension [15]. In the case of vowels, this transformation seems to be necessary to form an integrated phonetic representation.

The presence of bottom-up grouping cues characterizing vowel sounds, such as periodicity of the amplitude envelope, formant structure, and especially the combination of these features, leads to a rapid (within the first 80 ms) and long-lasting (~ 400 ms or longer) increase of the SPN in the auditory cortex [23,25,31]. The sensitivity of the SPN to bottom-up grouping signals and its ubiquitous presence in the neural response of children and adults suggests that it may serve as a candidate for detecting putative vowel processing dysfunction at the level of the auditory cortex in children with ASD.

Although previous EEG/MEG studies have not examined vowel-induced SPN in ASD individuals, many studies have focused on the sensor-level analysis of mismatch negativity/field (MMN/MMF) response to changes in vowels or syllables in the oddball paradigm. Despite considerable variation in results, their meta-analysis showed that individuals with ASD had increased MMN/MMF latencies and decreased MMN/MMF amplitudes in response to 'different phoneme' deviations, but not to phoneme-duration or phoneme-pitch deviations [32], suggesting specific disturbance of phonetic processing. The MMN/MMF studies, however, had certain limitations. *First*, they have not identified auditory cortical areas involved in abnormal vowel processing in ASD, either due to limitations of sensor-level analysis or coarse localization of brain activity based on a template brain. *Second*, MMN/MMF studies typically focused on the peaks of event-related responses, perhaps preventing detection of differences in activation beyond these peaks. *Third*, for reasons still under debate, transient ERP/ERF peaks in children and adults differ in timing, polarity, and underlying neural processes [33–35], making it difficult to compare data across age groups. *Fourth*, since the auditory cortex is sensitive to any regular acoustic patterns [16,17,19], atypical MMN may reflect a deficit that is not specific to the processing of speech sounds.

In the present study, we used a strategy that allowed us to clarify some of the previously unresolved questions regarding putative vowel processing deficit in ASD. *First*, we applied MEG imaging combined with individual brain models to localize cortical sources of atypical activity associated with analysis of vowels in children with ASD. *Second*, we tracked the entire timecourse of the vowel-related response in the auditory cortex rather than focusing on response maxima. *Third*, we focused on the SPN, which is present in both children and adults, overlapping with age-specific transient components of the auditory response, and may prove particularly informative for testing putative vowel processing deficits in the developing brain. *Fourth*, to test whether the deficit is specific to the 'speechness' of the vowel sound, we used synthetic vowels in which the formant structure and periodicity were either preserved or modified so that they could be analyzed separately.

We next asked whether atypical vowel processing is associated with impairment in the ability of children with ASD to distinguish speech from background noise. Such a possibility is worth investigating given the psychoacoustic evidence for the importance of vowels for speech intelligibility [36] and the evidence for impaired speech-in-noise recognition in ASD individuals with normal-range IQ and audiometrically normal hearing (for a review see [4]).

Although both words/syllables and sentences have previously been used to test speech-in-noise perception deficit in children with ASD, in the present study we decided to apply the 'words-in-noise' (WiN) paradigm. Words are less taxing on working memory than sentences, and their repetition is less dependent on cognitive ability, allowing this test to be used for verbal children with below-average intelligence. Our recent study showed that the ability to recognise two-syllable high-frequency words in background noise in verbal children with ASD does not correlate with their IQ [37]. Unlike the recognition of sentences, which carry higher-order linguistic information (overall tonal contour, stress pattern, syntactic structure, etc.), the recognition of isolated words relies more on bottom-up grouping cues. Thus, the words-in-noise paradigm is better suited than the sentences-in-noise paradigm to investigate the perceptual effects of a possible dysfunction in the relatively low-level cortical grouping processes underlying vowel identification.

To compare WiN perception in children with ASD and TD children, we used two types of background noises: stationary (ST) and amplitude-modulated (AM). Dips (i.e., short periods with more favorable SNR) in the AM noise allow listeners to improve speech perception. This improvement, referred to as "masking release", is attributed to the listener's ability to detect the target speech signal during the dips [38]. It has been suggested that atypically low 'masking release' in individuals with ASD is due to a reduced ability to integrate fragments of auditory information into meaningful words or sentences [39,40]. However, deficit in phonetic processing, especially of vowels, may also play a role, as the preserved spectral-temporal structure of vowels is particularly important for speech perception when speech is immersed in AM noise [36,38].

To summarize, we hypothesized that processing of natural auditory objects - vowels - is atypical in children with ASD and that this deficit may contribute to their degraded perception of words in noise. To test this hypothesis, we recruited children with ASD with a wide range of IQ and TD children and examined group differences in SPN using magnetoencephalography (MEG). To find out which vowel features processing is affected in ASD, we used synthetic periodic and aperiodic vowels, as well as complex periodic sounds that lacked vowel formant structure. We aimed to investigate when (in terms of processing timing) and where (in terms of auditory cortical regions) the timecourse of SPN in children with ASD diverges from the typical activation pattern. We then tested the relationship between atypical SPN activation patterns in children with ASD and their word recognition abilities in the ST and AM noise. Given the importance of vowels for 'dip listening', we

hypothesized that vowel processing deficit in children with ASD, if present, should predominantly affect their performance in AM noise.

MATERIALS AND METHODS

Participants

The study included 39 boys with ASD aged 6.9 – 13.0 years and 35 typically developing (TD) boys of the same age range. The TD children from the control group were recruited through advertisements in the media. None of the TD participants had known neurological or psychiatric disorders. Twenty two of 39 TD participants were the same as in our recent developmental study devoted to comparison of sustained negative responses to periodicity and formant composition of vowels in children and adults [25]. The participants with ASD were recruited through several sources (advertisements in the media, consulting centers, an educational center affiliated with the Moscow State University of Psychology and Education).

The ASD diagnosis was confirmed by an experienced psychiatrist who was the author of the current study (N.A.Y.) and was based on the Diagnostic and Statistical Manual of Mental Disorder (5th ed.) criteria as well as an interview with the child and parents/caregivers. In addition, parents of all children were asked to complete the Russian translation of parental questionnaires: Social Responsiveness Scale for children [41] and Social Communication Questionnaire (SCQ-Lifetime) [42] and most of them completed these questionnaires (Table 1). Intellectual abilities were assessed using the KABC-II test, and the Mental Processing Index (MPI) was used as an IQ equivalent [43]. This index excludes tasks that require verbal concepts, verbal reasoning, and cultural knowledge and is recommended for evaluation of non-verbal abilities in both TD children and those with ASD [44].

Hearing status of all participants was determined using pure tone air conduction audiometry with a AA-02 clinical audiometer ('Biomedilen'). Auditory sensitivity was tested at 500, 1000, 2000, 4000 Hz, and the threshold averaged over the four frequencies was calculated separately for each ear. All participants had normal hearing (threshold< 20 dB HL) in either ear [45].

The Ethical Committee of the Moscow State University of Psychology and Education approved this investigation. All children gave verbal consent to participate in the study and their caregivers gave written consent to participate.

Assessment of the general level of language development

The language test incorporated 12 subtests from the Russian Child Language Assessment Battery, RuCLAB [46], which evaluates expressive and receptive skills in vocabulary (production and comprehension of word), morphosyntax (sentence production and comprehension), and discourse (text production and comprehension). The audio material was provided by a professional female native Russian speaker and recorded in studio conditions. All test stimuli were delivered through the AutoRAT application [47]. The tests took place in a quiet and child-friendly environment. Before each subtest, participants received instructions and performed 2-3 practice trials, which were not included in the final analysis. The sequence of the test items was the same for all participants. During testing, the examiner completed a paper protocol. Additionally, the sessions were videotaped, adhering to ethical standards. The description of the test components and details on the scoring are provided in the Supplementary Methods and Table S1. To characterize the overall level of language proficiency, the arithmetic mean of scores on all subtests was calculated. This mean score is hereinafter referred to as the "total language score".

Words-in-Noise (WiN) test

Stimuli

The Words-in-Noise (WiN) test includes 160 two-syllable lemmatized Russian nouns with high imagery (ability to evoke mental images). These words were frequently occurring words according to the List of Frequent Nouns [48] and/or the list of 300 words most frequently used in daily life [49], and corresponded to the age norm for children 6 years and older. Two child psychologists who work professionally with children with ASD confirmed that these words are familiar to most verbal 7-year-old children with this diagnosis. The words were spoken by a 35-year-old woman with neutral, unemotional intonation and were recorded with studio recording equipment. The mean duration of the words was 694 ms (SD = 77 ms).

Mean loudness of each word was adjusted to correspond to that of 45 dB SPL pink noise using the "spl" function of a third-party software package for MATLAB R2020a [50]. The masking noise was of two types: stationary pink noise (ST) and amplitude modulated pink noise (AM). AM noise was obtained by modulating ST noise with a 10 Hz sinusoidal function (i.e., pink noise was interrupted 10 times per second); modulation started at the 0° phase. ST and AM noises were synthesized using 'pinknoise' and 'ammod' functions of MATLAB R2020a (MathWorks, Inc.). Power spectral density of the masking noise decreases proportional to its frequency (1/f), 3 dB per octave. Noise lasted for 1 second and started 75 ms before word onset. The rise/fall of each 1-second noise signal was smoothed at 1 ms intervals. Words were presented at a fixed sound pressure level. The sound pressure level of the masking noise was varied at four signal-to-noise ratios (SNR): -0, -3, -6, and -9 dB.

Testing procedure

The stimuli were presented through Sony WH-XB900N headphones with the noise-canceling function off. The headphones were calibrated using a CEM DT-815 sound level meter. A stationary pink noise of 45 dB was taken as the reference. The experiment began with a training session during which 10 words were presented randomly against a background of ST or AM noise (3 to -3 SNR). The participant was asked to repeat the word after each presentation. Only the exact repetition was considered a correct answer. The training lasted until the child internalized the instruction, but no more than 10 min. If the instruction was successfully learned, the experimenter proceeded to the main part of the test, which included four blocks in the sequence of 0, -3, -6, and -9 dB SNR, where the 0 SNR condition was the easiest and the -9 SNR condition was the most difficult. Each block contained 40 words - 20 words for each noise type (ST and AM). The type of the masking noise varied within a block in a pseudorandomized order (no more than three consecutive presentations of the same noise type). Words were selected randomly from the list and presented in one of the conditions. Each word was presented only once. For each condition (4 SNR levels, 2 types of masking noise), the number of correctly recognized words was counted. At the end of the main part of the experiment, words that the participant did not repeat correctly were presented without noise. All of the children in this study were able to repeat the words presented to them without noise. The WiN test was presented to 30 (of 39) TD children and to 29 (of 35) children with ASD. All TD and 27 of 29 ASD participants were able to complete the test.

Scoring

Although the words corresponded to the developmental level of the participants, some of them were particularly easy or difficult to recognise. We excluded from the analysis 18 words that were 'easiest' or 'most difficult' according to the accepted criteria (see Supplementary Methods). For each child, we

then calculated the average percent of correct responses in each of the SNRs (0dB, -3dB, -6dB, -9dB) in the ST and AM noise conditions. Many participants of both groups (14 of 30 TD and 14 of 27 ASD) gave 0% correct responses in the most difficult -9 dB ST noise condition. Therefore, we excluded the -9 dB SNR condition from further analysis. For each child, we then calculated the average percent of correct responses for each SNR level (0dB, -3dB, -6dB, -9dB) in the ST and AM noise conditions (WiNST and WiNAM scores, respectively).

MEG experiment

Stimuli

The experimental paradigm used in the present study is identical to that described in Orekhova et al. [25]. We used the four types of synthetic vowel-like stimuli previously used by Uppenkamp et al. [51] and downloaded from 'http://medi.uni-oldenburg.de/members/stefan/phonology_1/'. Five strong vowels were used: /a/ (caw, dawn), /e/ (ate, bait), /i/ (beat, peel), /o/ (coat, wrote) and /u/ (boot, pool).

The synthetic periodic vowels consisted of damped sinusoids, which were repeated with a period of 12 ms, so that the fundamental frequency of each vowel was 83.3 Hz. The carrier frequencies of each vowel were kept fixed at the four lower formant frequencies, which were chosen in a typical range of an adult male speaker. These are further referred to as periodic vowels. These regular vowel stimuli have been modified, as described below, to generate three other classes of stimuli: non-periodic vowels, periodic non-vowels, and non-periodic non-vowels. To violate periodicity, the start time of each damped sinusoid was jittered within ±6 ms relative to its start time in the original vowel, separately for each formant. Despite the degraded voice quality (hoarse voice), these non-periodic sounds were perceived as vowels. To violate formant constancy, the carrier frequency of each subsequent damped sinusoid was randomly chosen from a set of eight different formant frequencies used to produce regular vowels and randomized separately for each formant (Frequency range: formant 1 = 270 - 1300 Hz; formant 2 = 850 - 2260 Hz; formant 3 = 1750 - 3000 Hz; formant 4 = 3300 - 5500 Hz). Both periodic and non-periodic sounds with disrupted formant structure were perceived as noises rather than vowels (i.e. non-vowels). The following four stimulus types were presented during the experiment: (1) periodic vowels (/a/, /i/, /o/); (2) nonperiodic vowels (/a/, /u/, /e); (3) three variants of periodic non-vowels and (4) three variants of non-periodic non-vowels. The spectral composition of these stimuli is given in Supplementary Figure S1 (see also [25]).

Two hundred seventy stimuli of each of the four classes were presented, with three stimulus variants equally represented within each class (N = 90). All stimuli were presented in random order. Each stimulus lasted 812 ms, including rise/fall times of 10 ms each. The interstimulus intervals (ISIs) were randomly chosen from a range of 500 to 800 ms.

The non-periodic non-vowels were used as control stimuli. The contrasts of interest were (1) 'non-periodic vowels versus non-periodic non-vowels', (2) 'periodic non-vowels versus non-periodic non-vowels' and (3) 'periodic vowels versus non-periodic non-vowels'. By comparing these contrasts in the ASD and TD groups we investigated group differences in the processing of formant structure, periodicity/pitch, or a combination of these features in a natural vowel.

Procedure

Participants were instructed to watch a silent video (movie/cartoon) of their choice and ignore the auditory stimuli. Stimuli were delivered binaurally via plastic ear tubes inserted into the ear channels. The tubes were attached to the MEG helmet to minimize possible noise from their contact with the

subject's clothing. The intensity was set at 90 dB SPL. The experiment included three blocks of 360 trials, each block lasting around 9 minutes with short breaks between blocks. If necessary, the parent/caregiver remained with the child in the MEG shielded room during the recording session.

MRI data acquisition and processing

In all participants with ASD and in 28 TD participants T1-weighted 3D-MPRAGE structural image was acquired on a Siemens Magnetom Verio 3T scanner (Siemens Medical Systems, Erlangen, Germany) using the following parameters: [TR 1780 ms, TE 2.78 ms, TI 900 ms, FA 9°, FOV 256 × 256 mm, matrix 320 × 320, 0.8 mm isotropic voxels, 224 sagittal slices]. In 9 TD subjects MRIs were acquired at a 1.5T Philips Intera. In 2 TD subjects 1.5T GE Brivo MR355/MR360 was used. Cortical reconstructions and parcellations were generated using FreeSurfer v.7.4.1 [52,53].

MEG data acquisition, preprocessing and source localization

MEG data were recorded at the Moscow Center for Neuro-cognitive Research (MEG-Center) using Elekta VectorView Neuromag 306-channel MEG detector array (Helsinki, Finland) with 0.1 - 330 Hz filters and 1000 Hz sampling frequency. Bad channels were visually detected and labeled, after which the signal was preprocessed with MaxFilter software (v.2.2) in order to reduce external noise using the temporal signal-space separation method (tSSS) and to compensate for head movements by repositioning the head in each time point to an 'optimal' common position (head origin). This position was chosen individually for each participant as the one that yielded the smallest average shift across all data epochs after motion correction.

Further preprocessing steps were performed using MNE-Python software (v.1.4.1) [54]. The data were filtered using notch-filter at 50 and 100 Hz and a 110 Hz low-pass filter. Periods in which peak-to-peak signal amplitude exceeded the thresholds of 7e-12 T for magnetometers or 7e-10 T/m for gradiometers, and 'flat' segments in which signal amplitude was below 1e-15 T for magnetometers or 1e-13 T/m for gradiometers were automatically excluded from further processing. To correct cardiac and eye movement artifacts we recorded ECG, vEOG, and hEOG and used a signal-space projection (SSP) method. Next, we excluded from analysis data segments in which head rotation exceeded a threshold of 10 degrees/s along one of the three space axes, head motion velocity exceeded a threshold of 4 cm/s in 3D space, or head position deviated from the head origin position by more than 10 mm in 3D space. The data were then epoched from -0.2 s to 1 s relative to stimulus onset. The mean number of artifact-free data epochs was initially higher in the TD participants (TD: 997 vs ASD: 927, $p < 0.05$). To equalize this number, we randomly removed 70 epochs for each TD participant. The resulting mean number of clean epochs for each subject and stimulus type was 231 (range 141 - 340) and 231 (range 131 - 347) for TD and ASD children, respectively. The epoched data were averaged separately for the four experimental conditions and then baseline corrected by subtracting the mean amplitude in -200 - 0 ms prestimulus interval.

To obtain the source model, the cortical surfaces reconstructed with the Freesurfer were triangulated using dense meshes with about 130,000 vertices in each hemisphere. The cortical mesh was then resampled to a grid of 4098 vertices per hemisphere, corresponding to a distance of about 4.9 mm between adjacent source points on the cortical surface.

To compute the forward solution, we used a single layer boundary element model (inner skull). Source reconstruction of the event-related fields was performed using the standardized low-resolution brain electromagnetic tomography (sLORETA) [55]. Noise covariance was estimated in the time interval from -200 to 0 ms relative to stimulus onset. To facilitate comparison between

subjects, the individual sLORETA results were morphed to the fsaverage template brain provided by FreeSurfer.

MEG data analysis

Data analytic plan

First, to find out whether SPN associated with processing vowel periodicity and formant structure presents in both TD children and children with ASD, we compared responses to test stimuli (periodic vowels, non-periodic vowels, periodic non-vowels) and control stimuli (non-periodic non-vowels) separately in TD and ASD groups. This was done in the sensor space using a global root mean square RMS signal and then in the source space using the sLORETA values in combination with the nonparametric permutation test with spatiotemporal threshold-free cluster enhancement (TFCE) [56]. To analyze timing of the SPN associated with processing of periodicity or formant structure, we compared timecourses of neural current evoked by test and control stimuli in '*between-conditions* regions of interest' - *bc*ROIs. The *bc*ROI was chosen separately for each test-control stimulus contrast, based on the significance of the differences between test and control conditions.

Second, we used TFCE cluster analysis to test for the group differences in differential responses (e.g. response to periodic vowel minus response to control stimulus). Then, we analyzed the timing of the group differences in the SPN (i.e., negative differential response). To do so, we compared timecourses of differential responses in '*between-groups* regions of interest', *bg*ROIs, chosen on the basis of significance of the group differences in the SPN.

Third, when group differences in SPN were found, we tested the association between atypical SPN and WiN scores in children with ASD using partial correlation analysis to account for the effects of age and IQ.

Root mean square (RMS) analysis

For the sensor level analysis, all subjects' data were transformed to a common standard head position (to [0, 0, 45] mm, default MaxFilter parameter). RMS metric was computed based on the signal from all gradiometer sensors and compared between test (periodic vowels, periodic non-vowels, non-periodic vowels) and control (non-periodic non-vowel) conditions point-by-point in the 0 - 800 ms stimulation interval using paired t-tests.

Spatiotemporal cluster analysis

The spatiotemporal clustering analysis was performed in the 'labels' that were selected in the fsaverage template brain so as to broadly overlap the left and right auditory cortex and nearby areas (535 vertices per left hemisphere and 540 vertices per right hemisphere; Fig. 3). These labels were identical to those used in our previous study [25]. The direction of the source current in these labels was aligned using the MNE-python function 'label_sign_flip'. We then verified that in each participant, the timecourse averaged over all the point sources and across conditions had a negative sign between 300-800 ms, consistent with the sustained negativity observed in response to these stimuli in the auditory cortex [23,25] and a positive sign of the P100m component between 50-150 ms. Otherwise the signs of all point timecourses were flipped by applying a multiplication factor of '-1'. The source current data were cropped to 0-800 ms timeregion and downsampled to 500 Hz. Thus, the spatiotemporal clustering was performed with a total number of 214,000 (400 time points × 535 vertices) data points per left hemisphere and 216,000 (400 time points × 540 vertices) data points per right hemisphere.

TFCE spatiotemporal cluster analysis was used to compare evoked source currents between groups. The TFCE procedure derives the spatiotemporal cluster-level statistics for each data point p by using a weighted average between the cluster extent e (i.e., number of connected above-threshold data points) and the cluster height h (i.e., the statistical value of the data point), calculated according to the formula:

$$TFCE(p) = \int_{h=h_0}^{h_p} e(h)^E h^H dh,$$

where default values of $E = 0.5$ and $H = 2$ were applied [56], and starting threshold $h_0 = 0$, step size $dh = 0.4$ were set up for the better approximation. Permutation tests with TFCE effectively address the issue of multiple comparisons, and in this study, we computed 5000 TFCE permutations. For each data point, a corrected p-value was calculated by determining the proportion of permutations in which the TFCE output was greater than or equal to the original TFCE output. The null hypothesis of no cluster of the difference between the groups or conditions (within each group) was rejected at $p < 0.05$. As no interhemispheric spatial-adjacency-based clusters were expected, the 'check_disjoint=True' option was applied, which is equivalent to running the test separately for each hemisphere.

To identify brain regions showing significant differences between test and control stimuli we used one-sample permutation test with spatiotemporal TFCE ('stats.permutation_cluster_1samp_test' MNE Python function) separately in each group. In this test, the inputs were the differences between responses to test and control stimuli that were permuted within a single subject.

To identify clusters of significant Group × Condition interactions, we employed a two-sample permutation test with spatiotemporal TFCE ('stats.spatio_temporal_cluster_test' MNE Python function), where the inputs were the differences between responses to the test and control stimuli for each subject from the ASD and TD groups.

Analysis in the regions of interest (ROIs)

All ROIs for timecourse analyses were defined within the 'STG+' label that included auditory regions in the STG and posterior insular (A1, 52, LBelt, PBelt, MBelt, RI, A4, TA2, PI and PoI1, according to the HCPMMP1 atlas [57]). All these areas are structurally connected [58,59] and participate in processing of auditory information [60–62]. Although the activity in response to the auditory stimulation has also been observed in the superior segment of the circular insula (sINS) and in the superior temporal sulcus (STS), inspection of the polarity of the neural current suggests that in these areas it resulted from spreading activity from auditory cortical areas in the STG [63,64].

Vertex sources with the most significant differences, corresponding to the lowest TFCE p-values ('most significant' vertex sources), were chosen as ROIs. Specifically, for each vertex from the part of the TFCE cluster belonging to the STG+ label, we averaged the p-values over the cluster's time points and selected the vertices, whose averaged p-values were below the overall mean of the global average. A similar approach to select ROIs for timecourse analysis was applied by Alho et al. [65]. The spatial ROIs were defined separately for the left and right hemispheres and for each comparison performed (i.e., test-vs-control and ASD-vs-TD contrasts). Between-condition ROIs (*bc*ROIs) were formed based on the cluster of the differences in sLORETA values between test and control conditions, separately within each group (TD, ASD). The bcROIs were further used to refine time intervals when the differences between conditions were significant. Between-group ROIs

(*bg*ROIs) were formed based on the clusters of differences in differential responses (i.e., test minus control) between the TD and ASD groups. The bgROIs were further used to specify time intervals when the group differences in the SPN were significant.

To analyze correlations of the SPN with WiN scores in children with ASD, we first averaged *bg*ROI timecourses in the time interval where differences between TD and ASD groups were strongest. The temporal boundaries of the averaging were restricted to the mean onset and end times of significant group differences in SPN in the the *bg*ROI vertices (see Supplementary Figure 2). Using linear regression analysis, we then calculated the adjusted SPN (SPNadj) by regressing from the response to the test stimulus the magnitude of the response to the control (non-periodic non-vowels) stimulus and the square root of the mean number of averaged epochs per condition. Thus, by computing SPNadj, we accounted for large interindividual variability in the magnitude of auditory responses and for differences in the number of averaged epochs.

Statistical analysis

Nonparametric tests (Mann-Whitney U Test, Spearman correlations, or partial Spearman correlations) were used when the distribution differed significantly from normal according to Shapiro-Wilk's W test ($p < 0.05$). Otherwise parametric tests (T-test, Pearson correlations, Pearson partial correlations) were used. To calculate partial correlations we used the 'pcor' function in R. Group or condition-related differences in neural responses were assessed using TFCE cluster analysis (see above). In case of multiple comparisons, the False discovery rate (FDR) method of Benjamini and Hochberg [66] for correcting p values was applied. The accepted significance level was $p < 0.05$. To analyze the effect of SNR (0 dB, -3 dB and -6 dB) and type of masking noise (ST, AM) on the difference in WiN scores between the TD and ASD participants, we employed Mann-Whitney U Test and/or linear mixed model analysis using the 'lmer' function in R.

RESULTS

Characteristics of the samples

The characteristics of the samples are summarized in Table 1. Participants with ASD compared to TD participants had significantly higher SRS and SCQ scores. In the case of the SRS questionnaire, all but 3 of 34 tested subjects with ASD diagnosis had total scores above 60 points cut-off for ASD. For the SCQ questionnaire, all but 3 of 32 tested ASD subjects scored above the 15 points cut-off. The three ASD subjects who scored below the cut-off on the SRS questionnaire scored above the cut-off on the SCQ questionnaire, and vice versa. In the TD sample all but 5 children scored below ASD threshold on the SRS questionnaire, and all but one scored below ASD cut-off on the SCQ questionnaire. Given the sensitivity and specificity of the SRS and SCQ questionnaires [67,68], these results suggest that our samples of ASD and TD children are very well separated on autistic traits.

The participants with ASD as compared with TD participants had significantly lower MPI IQ scores. 18 of 34 ASD subjects for whom the results of the KABC-II test were available had the MPI IQ scores below 85, which corresponds to 'below-average' intelligence [43].

All, but two children with ASD displayed a history of language delay or impairment, defined by parents' reports of lack of two-word combinations at age three or the presence of language problems at the time of the diagnostic assessment. Six children with ASD experienced language regression at some point in their development.

Table 1. Characteristics of the samples.

	Mean (range) ± SD or Median (range)		Student T-test / M-W U test*, Z	p
	TD	*ASD*		
Age	Mean = 9.8 ± 1.8 (7.3 - 13.0) N = 39	Mean = 10.2 ± 1.7 (6.9 - 12.8) N = 35	T = 0.87	0.39
MPI IQ (KABC-II)	Mean = 116.2 ± 13.3 (85 - 138) N = 39	Mean = 83.1 ± 15.8 (54 - 128) N = 34	T = 9.7	< 0.0001
SRS	Mean = 37.6 ± 20.0 (3 - 84) N = 37	Mean = 99.5 ± 27.2 (33 - 157) N = 34	T = 11.0	< 0.0001
SCQ	Mean = 6.3 ± 3.2 (2 - 16) N = 38	Mean = 24.8 ± 6.4 (12 - 35) N = 32	T = 15.6	< 0.0001
Language total score**	Median = 94.9 (86.3 - 98.6) N = 29	Median = 79.7 (35.4 - 94.1) N = 32	Z = 6.2	< 0.0001

N - number of participants for whom information is available.
* Mann-Whitney U Test was applied for the group comparison in Language total scores, because their distribution in the ASD group was significantly different from normal.
** language proficiency was estimated using the RuCLAB test.

Words in noise (WIN) test results in TD and ASD groups

Decreasing SNR worsened performance in both groups of participants and for both types of noise ('lmer' with subject and noise-type as random factor, for 0 dB vs. -6 dB comparison: Ts > 7.8, ps < 3e-10 for both ST and AM noise). Children with ASD recognised significantly fewer words than TD children in AM noise at all SNRs and in ST noise at -3 and -6 dB SNR (Fig. 1A). Both TD and ASD children performed better in case of the AM than ST noise, with the exception of the 0 dB SNR condition in the ASD group, when no 'masking release' was observed (Fig. 1B). We employed linear mixed model analysis to investigate whether the effect of Noise Type (ST, AM) differed between TD and ASD groups. Subject and SNR (0 dB, -3dB and -6dB) were taken as random factors. There was a trend for Group x Noise Type interaction ($t(276) = 1.83$, $p = 0.068$), suggesting that the facilitating effect of the amplitude-modulated noise might be lower in ASD than in TD participants (Fig. 1B). Given this trend, as well as previous literature showing different masking properties of the ST and AM noise [39,40], for each child we calculated WiNst and WiNam scores for ST and AM noise separately, as the percentage of correct responses averaged over SNR levels.

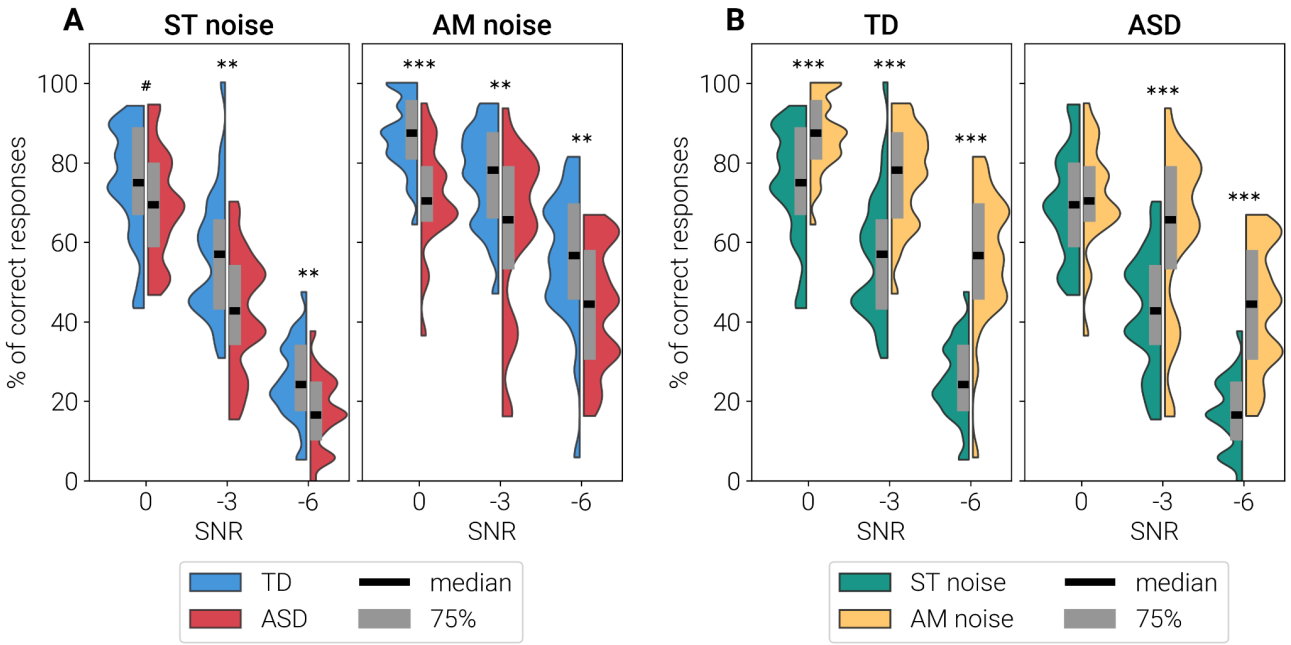

Fig. 1 - Mean percentage of correctly repeated words presented against a background of masking noise. **A:** Comparison of performance in TD and ASD groups, separately for the stationary and amplitude-modulated noise conditions. **B:** Comparison of performance in stationary and amplitude-modulated noise, separately for TD children and children with ASD.
# p < 0.1, ** p < 0.01, *** p < 0.001 (Mann-Whitney U Test).

In both groups, WiNST and WiNAM scores improved with age (Spearman correlations; WiNST: TD, R(N=30) = 0.43, p = 0.02; ASD R(N=26) = 0.50, p = 0.009; WiNAM: TD R(N=30) = 0.56, p = 0.001, ASD R(N=26) = 0.51, p = 0.008). WiN scores did not correlate with IQ in either group (Spearman correlations; WiNST: TD R(N=30) = 0.08, p = 0.65, ASD R(N=26) = -0.27, p = 0.18; WINAM: TD R(N=30) = 0.10, p = 0.58, ASD R(N=26) = 0.02, p = 0.93).

To sum up, the ability to recognize words in noise was impaired in children with ASD compared to their TD peers, was independent of IQ, and improved with age.

WiN performance and general language abilities in TD and ASD groups

As expected, children with ASD had significantly lower total language scores than TD participants (Mann-Whitney U Test: Ntd = 29, Nasd = 32, Z = 6.2, p < 0.0001, $\eta 2$ = 0.64). The scores improved with age in control participants (Spearman R (N=29) = 0.74, p < 0.0001), but not in children with ASD (Spearman R (N=32) = 0.21, p = 0.23). Unlike WiN scores, language total scores correlated with IQ (Spearman correlations, TD (N=29): R = 0.35, p = 0.06; ASD (N=31): R = 0.46, p = 0.008).

We further calculated in children with ASD Spearman partial correlations between language total score and WiN scores while controlling for age and MPI IQ. There was a tendency for partial correlations between language total score and WiNam (R*part* (N=24) = 0.40, p = 0.06). For WiNst the correlation was in the same direction though also not significant (R*part* (N=24) = 0.29, p = 0.19). The correlation with language total score became significant for the average of WiNst and WiNam scores (R*part* (N=24) = 0.43, p = 0.048). Thus, the ability of children with ASD to correctly recognize words presented in noise may contribute to their language skills regardless of age and nonverbal IQ.

MEG results: sensor-level analysis

Figure 2 shows the grand average auditory evoked field waveforms expressed as root mean square (RMS) signals calculated over all 204 gradiometer channels. Compared to control stimuli, stimuli

characterized by temporal regularity (periodic non-vowels), formant structure (non-periodic vowels), or a combination of these features (periodic vowels) caused a transient decrease in RMS around 100 ms relative to post stimulus onset (time range of the child P100m component), followed by its prolonged increase after ~ 150 ms that lasted up to 400 ms or longer. The RMS amplitude provides a measure of the magnetic field strength across the MEG sensors regardless of polarity, and is blind to the true direction of the condition-related difference in the cortical currents [69]. In our previous study we have shown that both the decreases and increases in RMS in response to vowel-like versus control stimuli were explained by sustained negative shift of current in neural activity of the superior temporal cortex [25], i.e., SPN. To account for current polarity, we further analyzed the data at the cortical source level.

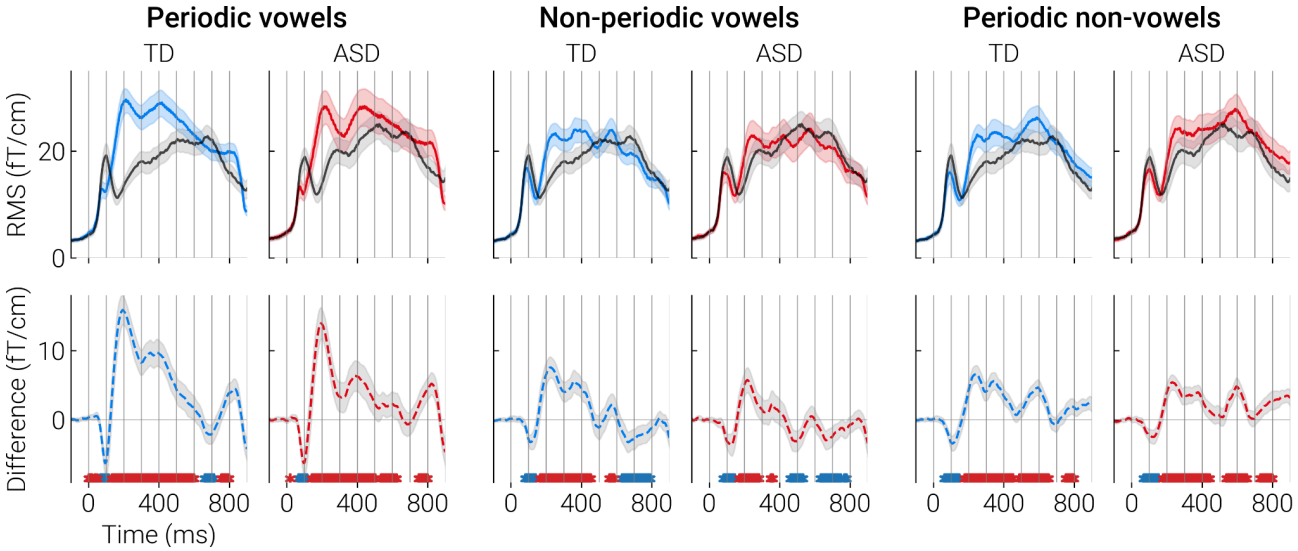

Fig. 2 - Grand average RMS response waveforms and the RMS difference waves in ASD and TD groups. Zero point at the horizontal axis corresponds to the onset of 800-ms stimulus. RMS was calculated over all gradiometer channels. Black line denotes control condition (non-periodic non-vowels) and coloured lines denote test conditions (periodic vowels, non-periodic vowels, periodic non-vowels). Dashed lines indicate differential responses. The asterisks under the dashed lines correspond to significant point-by-point differences between the test and control conditions: red color - control < test; blue color - control > test (paired t-test, $p < 0.05$, FDR corrected). The shaded areas mark 95% confidence intervals.

MEG results: analysis of the source current

Effects of periodicity and formant structure in TD and ASD participants

To test whether children with ASD, similarly to TD children [25], respond to patterned acoustic input with a SPN in the auditory cortical areas, we applied spatiotemporal cluster analysis separately in TD and ASD groups. SPN was defined as an increase in negativity in the evoked source current for the Test versus Control contrasts, i.e., a negative sign of the difference between vowel or complex periodic sounds (periodic non-vowels) and a control non-periodic non-vowel sound.

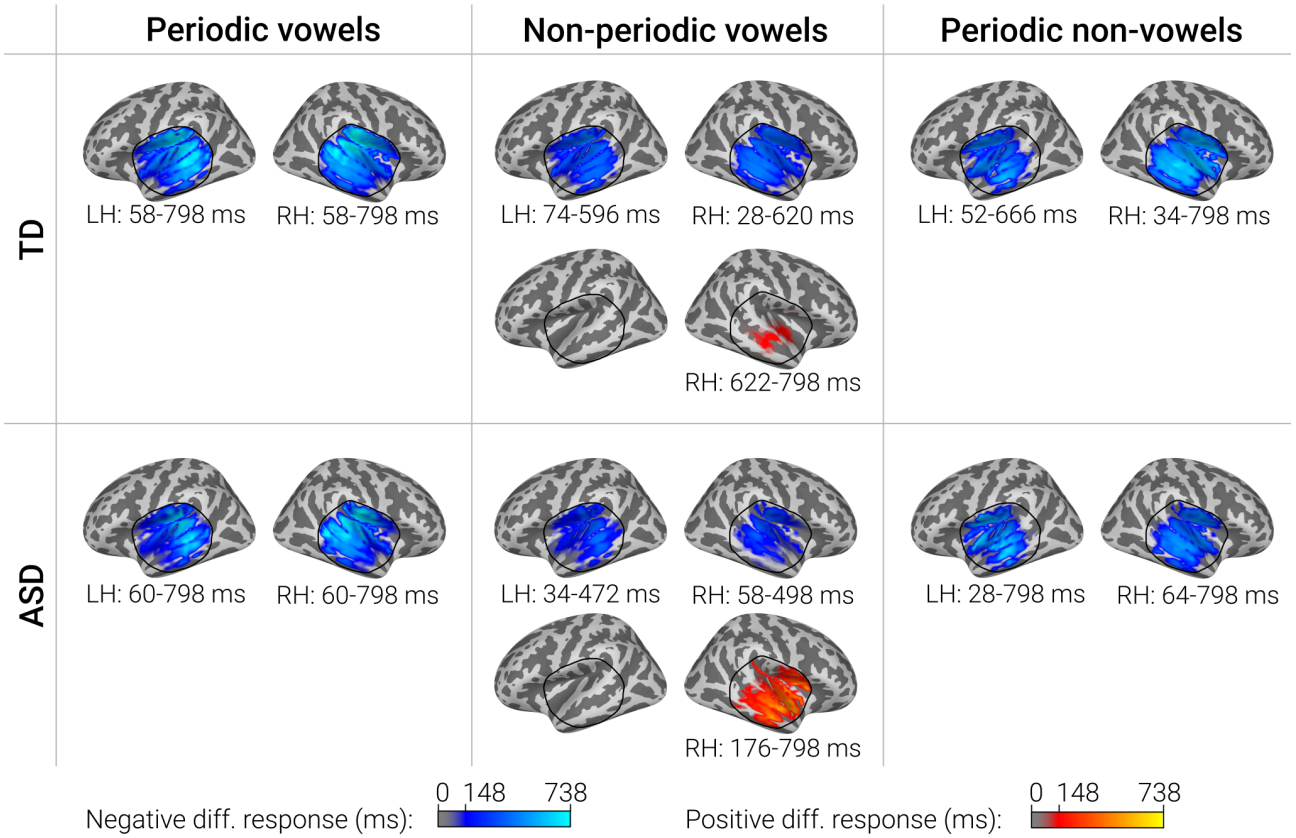

Fig. 3 - Significant clusters of the differences in the evoked source current between test (periodic vowel, non-periodic vowel, periodic non-vowel) and control (nonperiodic non-vowel) conditions in TD and ASD groups. Blue colors correspond to SPN, i.e. a more negative source current to the test compared to the control condition, red colors - to the opposite direction of the difference. Color intensity indicates the duration of the cluster. Black line indicates the border of the region used for cluster analysis. Note that the 798 ms point is the last time point analyzed.

Cluster analysis was performed in the labels broadly overlapping the auditory cortex and nearby regions where auditory evoked activity was observed (Fig. 3). Before cluster analysis, the direction of dipole sources within this label was adjusted to correspond to the dominant direction of source current in the auditory cortex (see Methods for details). In both groups, periodicity, formant structure, and especially their combination were associated with bilateral clusters of negative differential responses, i.e., SPN that lasted several hundred milliseconds (Fig. 3). The SPN clusters encompassed the primary and secondary auditory cortex, as well as areas in the superior temporal sulcus and medial temporal gyrus. In case of non-periodic vowels, the cluster of negative differences (test > control) was followed by (or co-existed with) a cluster of positive differences (test < control). In our previous study a similar pattern of differences between non-periodic vowel and control stimuli was observed in neurotypical adults (see [25] for discussion). The temporal evolutions of the clusters of differences between test and control stimuli in the TD and ASD groups are shown in Supplementary videos S1-S6.

To examine the SPN timing, we compared timecourses of the source current evoked by test and control stimuli in the *bc*ROI (ROI where the differences *between conditions* were most significant; see Methods). Figure 4 illustrates the *bc*ROIs for the periodic vowels and the responses in these bcROIs (Fig. 4 A, B and C, D). The differential responses (periodic vowels minus control stimuli) and their significance (i.e. difference from zero) are shown in Fig. 4 E, F and Fig. 4 G, H, respectively. Similar illustrations for non-periodic vowels and periodic non-vowels are given in the Supplementary materials (Figures S3 and S4).

In both groups and in response to all test stimuli, significant SPN (i.e. negative sign of the difference in the source current between responses to test and control stimuli) began earlier than 100 ms after stimulus onset and then persisted for approximately 500 ms. This timing is consistent with previous results indicating early neural discrimination of auditory patterns characterized by some grouping cues [16,70], which in our study were represented by frequency composition and/or periodicity. The group differences in the SPN are described in the next section.

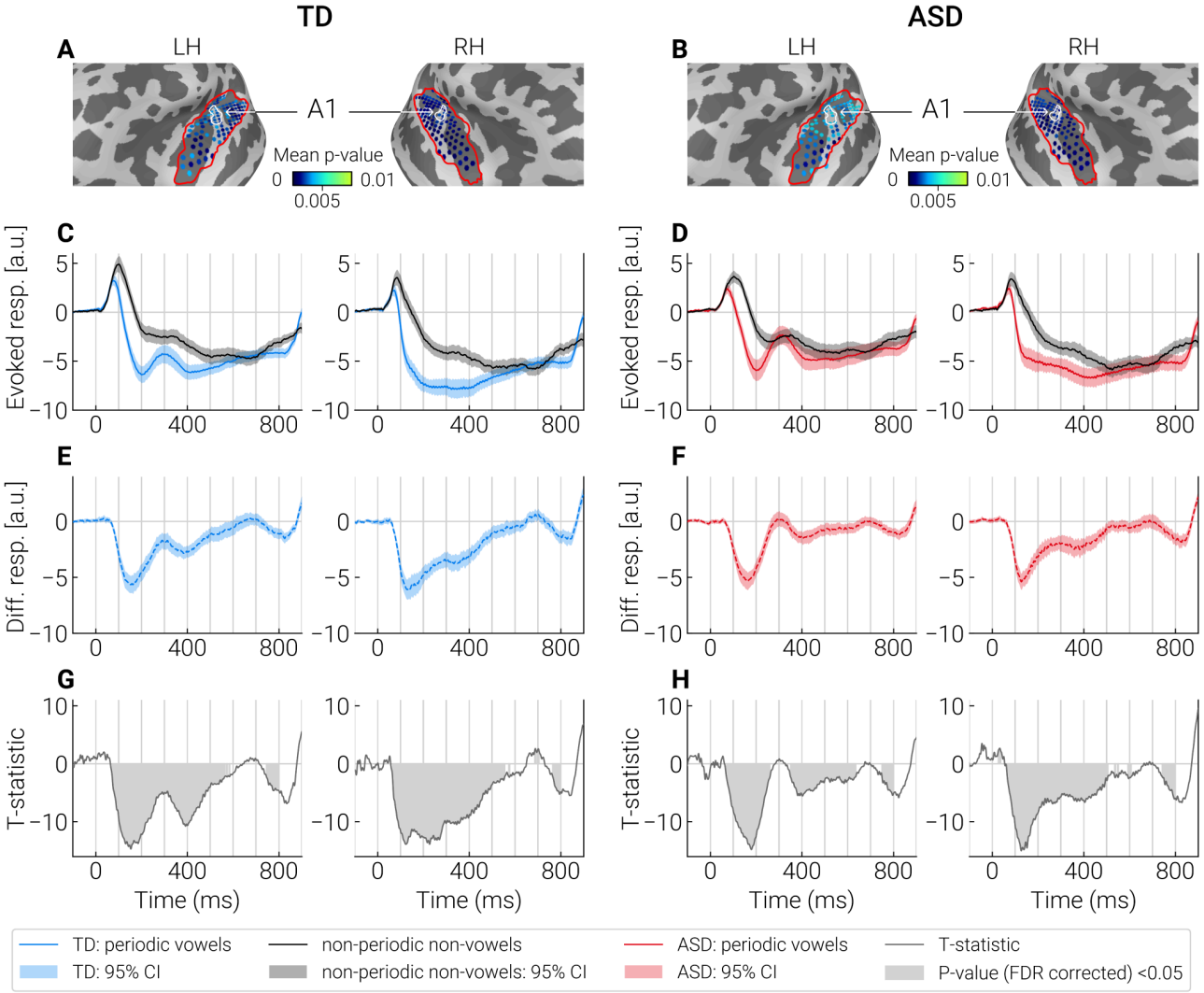

Fig. 4 - Comparison of evoked responses to the periodic vowels and control stimuli (non-periodic non-vowels) in the *bc*ROIs. **A, B:** Red line outlines the label used for ROI selection. Within the labels, the 'most significant' dipole sources representing the *bc*ROIs are marked by blue dots. Color shade (light blue to dark blue) indicates significance of the differences between test and control conditions in the respective point sources belonging to *bc*ROIs. The primary auditory cortex (A1) is outlined with a white contour. **C, D:** Averaged neural current timecourses in the *bc*ROIs. **E, F:** The difference between timecourses of current evoked by the test and control stimuli in the *bc*ROIs. **G, H:** T-statistics reflecting a pointwise comparison of the response timecourses to test and control stimuli. Significant differences ($p < 0.05$, FDR corrected) are marked in gray. Supplementary figures S3 and S4 show the similar results for non-periodic vowels and periodic non-vowels.

Group differences in SPN related to the processing of periodicity and formant structure

To compare SPN between TD and ASD groups we ran TFCE cluster analysis for the differential responces (differences between responses to test and control stimuli). The analyses yielded bilateral clusters of the group differences for differential responses to non-periodic and periodic vowels (i.e. sounds characterized by the presence of formant structure), but not for periodic non-vowels. The spatial location of the clusters on the surface of the 'inflated' brain is shown in Figure 5.

Supplementary Figure 6 shows the same clusters projected onto the surface of white matter for a three-dimensional representation.

In all cases, group differences were driven by greater SPN in TD compared to ASD group. As shown in Figure 5, the cortical localization of clusters of significant group differences was remarkably similar for periodic and non-periodic vowels. Within the STG+ labels, vertex sources with the highest significance and greatest temporal extent of the group differences were located anteriorly and laterally or medially relative to the primary auditory cortex (area A1). In the left hemisphere, the most significant and long-lasting differences were observed in the anterior part of the parabelt auditory area A4 [57]. In the right hemisphere, the differences were most prominent in the posterior segment of the circular insula sulcus (pINS). Additionally, the group differences in SPN were localized to the superior segment of the circular insula sulcus (sINS) and STS. However, inspection of direction of the current (Figure S5) and of the clusters' projection on the surface of the white matter (Figure S6) suggest that the differences in these regions are likely the result of point spread from the sources belonging to auditory areas [63,64].

To sum up, neural processing of sounds with vowel frequency composition differed between children with ASD and TD children in the non-primary auditory cortex of both hemispheres. These group differences in the SPN had a more restricted localization than the SPN itself (see Fig. 4). SPN associated with periodicity of non-vocal sound did not differ between the groups.

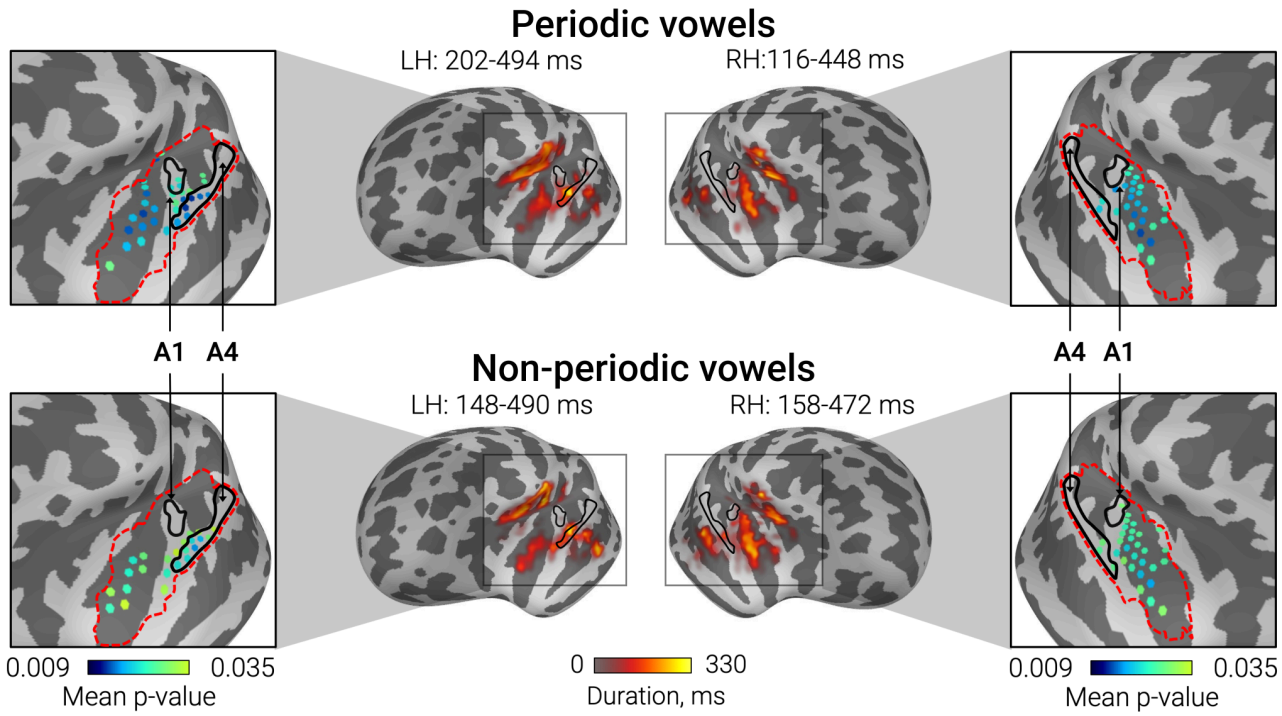

Fig. 5 - Differences between TD and ASD groups in sustained processing negativity (SPN) associated with periodic and non-periodic vowels. Central panels show the cortical localization of the TFCE clusters of significant group differences in differential responses (vowels minus control non-periodic non-vowel stimulus). Color bar below the inflated surfaces indicates temporal extent for sources belonging to the TFCE cluster. The left and right panels show the 'most significant' vertex sources selected based on the probability of the SPN group differences within the STG+ label delineated by the red dashed line. Color bars below the images indicate vertices' p-values averaged over the temporal extent of the cluster. These sets of vertices were used as *bg*ROI to analyze the timecourses of SPN group differences.

Current timecourses associated with processing of vowel formant structure: SPN and P3a-like response

To analyze timecourses of the group differences in SPN, we averaged auditory responses in the *bg*ROIs, i.e., in the regions where these group differences were most significant (Fig. 6). The group differences emerged starting from ~ 150 - 200 ms after stimulus onset, and persisted for approximately 200 ms (Fig. 6. E, F).

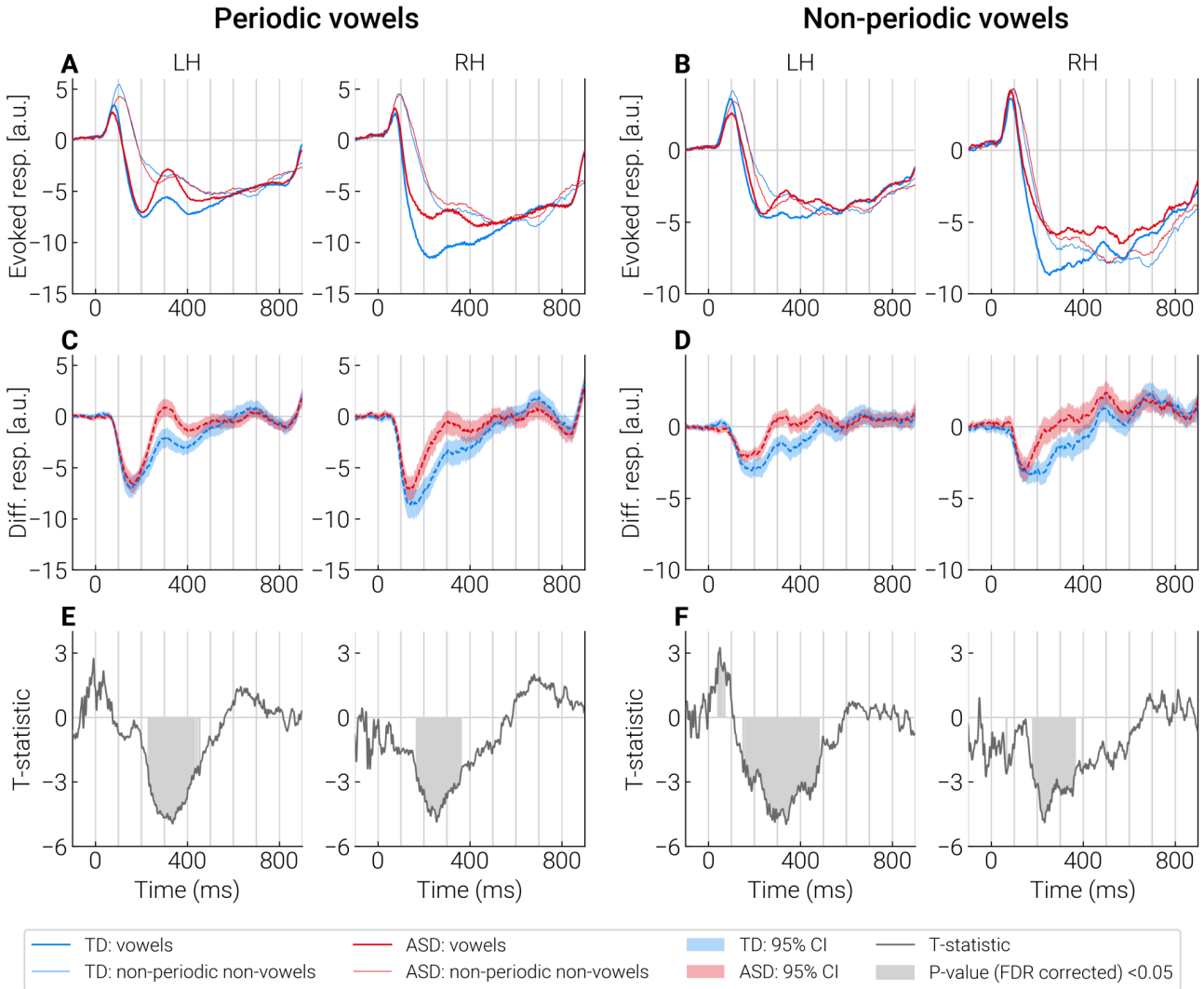

Fig. 6 - Timecourses of evoked responses in the auditory *bg*ROIs discriminating between TD and ASD groups. **A, B:** Group average timecourses of responses to vowels and control stimuli in TD and ASD groups. **C, D:** Average timecourses of the differential responses (test stimulus - control stimulus) in TD and ASD groups. **E, F:** Point-by-point comparison of timecourses of differential responses between the groups. Significant differences ($p < 0.05$, FDR corrected) are marked in gray.

To ensure that the differences in the SPN between the groups were due to test rather than control stimuli, we compared responses to these stimuli in TD children and children with ASD. The responses were preliminary averaged across a *bg*ROI in the interval bounded by the mean onset and end times of significant group differences (see Methods and Supplementary Figure S2). No differences were found in responses to control stimuli (all p's > 0.3). The responses to periodic and non-periodic vowels were reduced in children with ASD compared to TD children (periodic vowel, left hemisphere: $t(72) = 3.02$ $p = 0.003$, Cohen's $d = 0.67$, right hemisphere: $t(72) = 4.44$, $p = 0.00003$, Cohen's $d = 0.92$; non-periodic, left hemisphere: $t(72) = 1.92$, $p = 0.06$, Cohen's $d = 0.44$, right hemisphere: $t(72) = 3.59$, $p = 0.0006$, Cohen's $d = 0.78$).

Inspection of Figure 6 revealed a positive transient peak with a latency of ~ 300 ms that was most prominent in response to periodic vowels (and less so to non-periodic vowels) in the left-hemispheric *bg*ROI. Considering recent intracranial evidence for left-hemispheric predominance of the P3a response associated with processing of novel and potentially salient speech sounds [71], this deflection might reflect an involuntary capture of attention to perceptually salient vowels presented in a sequence of less salient stimuli.

Group differences in this P3a-like response might obscure (or produce) group differences in SPN. To test this possibility, we performed an additional analysis. For each subject and each ROI separately, we low-passed the timecourse signal at 10 Hz and estimated P3a amplitude as the amplitude of the largest positive peak in the 200-400 ms range relative to the mean of the two nearest negative peaks. If a positive peak between 200 and 400 ms was absent or indistinct (i.e., its amplitude relative to the preceding or following negative peaks, was below the RMS of the signal amplitude in the baseline period), the P3a peak was considered absent and its amplitude was set to zero.

In the pooled sample of participants, the P3a in response to non-periodic vowels was more frequently detected in the left than in the right hemisphere (69% and 50% respectively; Chi square = 4.9, p = 0.02). In the left hemisphere, it was more frequently detected in response to nonperiodic vowels than control stimuli (69% and 45% respectively; Chi square = 8.64, p = 0.003), while no such difference was found in the right hemisphere (50% vs 42%; Chi square = 0.95, p = 0.33). There were no group differences in the occurrence of P3a in response to non-periodic vowels either in the left (TD: 64%, ASD: 74%; Chi square = 0.85, p = 0.36) or right (TD: 44%, ASD: 57%; Chi square = 1.2, p = 0.27) hemispheres, and its amplitude did not differ between the groups (Mann-Whitney U-test, left hemisphere: Z = 1.53, p = 0.12, $\eta 2$ = 0.03; right hemisphere: Z = 1.29, p = 0.17, $\eta 2$ = 0.02).

In response to periodic vowels, the P3a also occurred more frequently in the left than in the right hemisphere in the combined sample of participants (82% and 66% respectively; Chi square = 4.89, p = 0.03) and was more often present in response to periodic vowels than control stimuli (left: 82% and 46% respectively; Chi square = 0.85, p < 0.0001; right: 66% and 43%; Chi square = 7.8, p = 0.005). In response to periodic vowels, P3a in the left hemisphere was more frequently detected in the ASD than in TD group (TD: 72%, ASD: 94%; Chi square = 6.07, p = 0.014), whereas no group differences were found in the right hemisphere (TD: 67%, ASD: 66%; Chi square = 0.01, p = 0.93). Amplitude of the P3a in response to periodic vowels was higher in children with ASD than in TD children in the left hemisphere (Mann-Whitney U-test: Z = 2.93, p = 0.003, $\eta 2$ = 0.12), but not in the right one (Mann-Whitney U-test: Z = 0.67, p = 0.50, $\eta 2$ = 0.01).

The amplitude of the left-hemispheric P3a response to periodic vowels decreased with age in participants with ASD (Spearman R = - 0.45, p = 0.007), but not in TD participants (Spearman R = - 0.11, p = 0.52), although group differences in correlation coefficients did not reach the level of significance (Fisher's Z = 1.54, p = 0.13). No correlations with age were found for the left-hemispheric P3a-like responses to non-periodic vowels (TD: R = - 0.12, ASD: R = 0.03, n.s.).

To sum up, the SPN to vowel formant structure was reduced in children with ASD in the STG areas anterior and lateral to the A1 and in pINS. This reduction was observed in both hemispheres from ~ 150 - 200 ms to ~ 350 - 450 ms after stimulus onset. In the left hemisphere, periodic and non-periodic vowels generated P3a-like deflection that was superimposed on the sustained negativity. In response to periodic vowels, the left-hemispheric P3a was present more frequently and with larger amplitude in the ASD group than in the TD group, whereas no group difference was found for the non-periodic vowels. These findings indicated that the P3a response might contribute to diminished SPN to periodic vowels in children with ASD in the left hemisphere, but was unlikely to

account for the significant reduction of the SPN associated with processing of non-periodic vowels. In the right-hemisphere, P3a response can hardly explain the group differences in SPN to either periodic or non-periodic vowels.

SPN to vowels predicts WiN scores in children with ASD

To test whether atypical SPN to vowels in children with ASD predicts their WiN scores, we applied partial correlation analysis, controlling for age and IQ. SPN was preliminary adjusted (SPNadj) for the magnitude of response to control stimulus and for the number of averaged trials in the following two steps. *First,* for each subject, the timecourses of responses to vowel and control stimuli were averaged within *bg*ROI in an time interval bounded by the mean start and end times of significant group differences (see Methods and Supplementary Figure S2). *Second*, we computed SPNadj as regression residuals after partialling out the magnitude of response to control stimulus and the square root of the number of averaged epochs from the magnitude of response to the test stimulus.

The partial correlations of SPNadj with WiN scores are presented in Table 2, separately for periodic and non-periodic vowels. For both types of vowels, greater SPNadj in the right hemisphere was associated with better WiN scores in the AM noise (WiNam). Figure 7 illustrates the relationships between adjusted SPN to periodic and non-periodic vowels in the left hemisphere and WiNam performance in the ASD sample. No significant correlations were found for the WiNam in the right hemisphere or for the WiNst in either hemisphere.

Table 2. Relationship between psychophysical and MEG results: Spearman partial correlation between WiN scores and SPNadj to periodic and non-periodic vowels.

	Periodic vowels	**Non-periodic vowels**
	Left hemisphere	
WiNam	**SPNadj: R*part* = - 0.55, p = 0.005*** Age: R*part* = 0.50, p = 0.013* IQ: R*part* = 0.19, p = 0.36	**SPNadj: R*part* = - 0.61, p = 0.001*** Age: R*part* = 0.48, p = 0.02* IQ: R *part*= 0.13, p = 0.53
WiNst	SPN(MEG): R*part* = - 0.26 p = 0.22 Age: R*part* = 0.46, p = 0.02* IQ: R*part* = - 0.21, p = 0.32	SPN(MEG): R*part* = - 0.19, p = 037 Age: R*part* = 0.45, p = 0.02* IQ: R*part* = -0.23, p = 0.27
	Right hemisphere	
WiNam	SPN(MEG): R*part* = - 0.03, p = 0.88 Age: R*part* = 0.52, p = 0.01* IQ: R*part* = 0.10 p = 0.64	SPN(MEG):R*part* = - 0.01, p = 0.96 Age: R*part* = 0.51, p = 0.01* IQ: R*part* = 0.11 p = 0.62
WiNst	SPN(MEG):R = - 0.02, p = 0.92 Age: R*part* = 0.49, p = 0.01* IQ: R*part* = - 0.23 p = 0.28	SPN(MEG):R = - 0.02, p = 0.91 Age: R*part* = 0.49, p = 0.01* IQ: R*part* = - 0.23 p = 0.28

To assess the specificity of the correlations for WiNam, we compared coefficients of correlations of the left-hemispheric SPNadj with the WiNam and WiNst scores using Fisher's Z statistics for dependent samples [72]. In the case of non-periodic vowels, the correlation was significantly higher for WiNam than WiNst scores (N = 26, Z = 2.35; p = 0.018). No significant difference was found for the periodic vowels (N = 26, Z = 1.60; p = 0.11).

The WiN data were available for 26 of 35 children with ASD. WiN - words-in-noise. SPNadj - Sound Processing Negativity, adjusted for response to control stimuli and for the number of averaged epochs. Significant Spearman partial correlations of interest which survive Bonferroni correction for multiple comparison (uncorrected $p < 0.00625$) are marked in bold. P-values are two-tailed, uncorrected for multiple comparisons. * $p < 0.05$

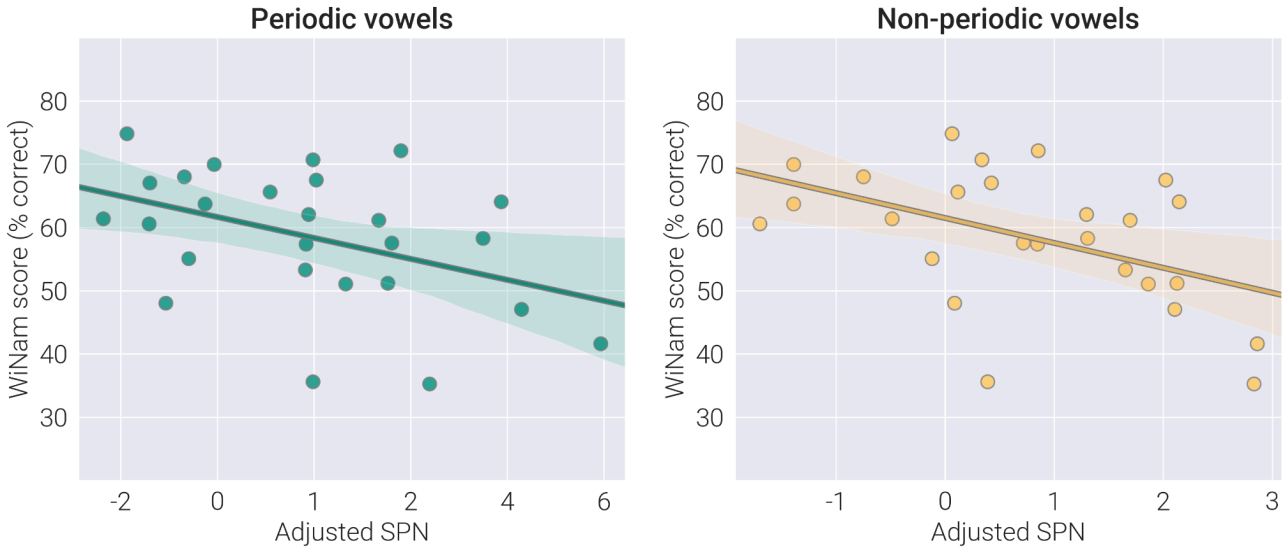

Fig. 7 - The relationship between WiNam scores and adjusted SPN to vowel stimuli in children with ASD. WiNam scores - percent of correctly repeated words in the amplitude modulated noise. Adjusted SPN - magnitude of the sustained processing negativity adjusted for the magnitude of response to control stimuli and for the number of averaged epochs. WiNam scores were corrected for age.

In the left hemisphere, SPN to periodic and non-periodic vowels overlapped with the P3a-like response, which could contribute to the correlation with the WiNam scores. To test this contribution we repeated the partial correlation analysis including P3a amplitude as an independent variable (Supplementary material, Table S2). The results suggest that the link between SPNadj and WiNam scores cannot be explained by the differences in the amplitude of the P3a-like component.

Children in the ASD sample had variable levels of IQ, which ranged from moderate intellectual disability to age-appropriate intellectual capacities (Table 1). Therefore, to evaluate the nonspecific effect of disorder severity, we tested whether the reduced SPNadj to vowels in children with ASD was related to their lower IQ scores. No significant Spearman correlations were found (periodic vowel, left hemisphere: R = - 0.03, n.s.; periodic vowel, right hemisphere: R = - 0.21, n.s.; non-periodic vowel, left hemisphere: R = 0.01, n.s.; non-periodic vowel, right hemisphere: R= - 0.2, n.s.).

DISCUSSION

We investigated whether cortical processing of isolated sounds characterized by periodicity and/or formant structure differs in the auditory cortex of children with autism and TD children, and whether differences related to the processing of these key vowel features contribute to poor perception of words in noise in children with ASD. In both groups of children presence of periodicity and formant structure was associated with sustained processing negativity (SPN) – an early (starting before 100 ms) negative shift of current in the primary and non-primary auditory cortical areas that persisted for several hundred milliseconds. We found no evidence for atypical periodicity (*f0*) processing of non-vocal spectrally complex sounds lacking formant composition in children with ASD. In contrast, the SPN evoked by vowel-like sounds carrying formants, was significantly reduced in ASD as compared with TD children, regardless of the periodicity of the sound. This SPN attenuation

emerged relatively late (around 150 -200 ms after a vowel onset) and was localized bilaterally to the auditory areas located anteriorly to the primary auditory cortex (parabelt area A4 and/or the pINS cortex). In the left, but not in the right hemisphere, reduced SPN in response to vowels predicted poor recognition of words presented against AM noise. Overall, our results suggest that impaired processing of vowel formant composition in children with ASD contributes to their impaired ability to recover words from glimpses of speech interrupted by masking noise.

Attenuated processing of vowel formant structure in children with ASD

The presence of SPN in response to periodicity or formant structure is consistent with the previous findings in neurotypical children and adults [23,25,70] and extends these findings to children with ASD, at least those with phrasal speech. Similar sustained negative shifts of current were observed in several recent MEG and EEG studies in response to sounds that can be characterized as acoustic patterns distinguished on the basis of their temporal properties such as periodicity [73] or frequency composition, either static or coherently changing in time [16,22,24,74–76]. It has been suggested that this negative shift reflects the fundamental cortical mechanism of automatic grouping in the auditory modality [76]. The early (< 100 ms) latency of the SPN in our study (Figure 4) is consistent with evidence for remarkably early sensitivity of the human auditory cortex to acoustic patterns [16] and, in particular, to vowels [70].

Being well-recognized vocal sounds deeply shaped by the experience of verbal communication, vowels are unique auditory objects. Studies have repeatedly shown that certain areas of the secondary auditory cortex and adjacent regions show a preference for conspecific vocalizations in humans [77–79] and non-human primates [61]. Although still debated [78], it has been suggested that voices are similar to faces in many ways, as both are 'special', carry information about the personality and emotional state of the subject, and are processed in specialized cortical areas [77].

In this respect, the decrease of SPN evoked by periodic and non-periodic vowel-like sounds in children with ASD is a remarkable finding. This decrease can be attributed specifically to an attenuated response to formant composition rather than to the periodicity of the vowel amplitude envelope (fundamental frequency / pitch), as the latter auditory cue was absent in non-periodic vowels. Since children with ASD had normal SPNs to nonvocal sounds characterized by f0 periodicity, as well as normal responses to control nonperiodic nonvocal stimuli (see Fig. 6 A and B), the reduced SPNs to vowels cannot be explained by a general decrease in response amplitude or non-specific deficit in auditory pattern processing.

Despite the early start of the SPN (< 100 ms post-stimulus onset), its group differences emerged relatively late (> 150 ms post-stimulus onset) and were located predominantly in the non-primary auditory areas (Fig. 5). Spared functional activity of the primary auditory cortex in children with ASD in our study is in line with fMRI findings in ASD individuals [80,81]. This result is also consistent with the results of Engineer et al. [82] who found in a mice model of autism that non-primary auditory cortical areas are more vulnerable to prenatal factors leading to autism than the primary auditory cortex.

Presence of typical SPN up to at least 150 ms after vowel onset (Fig. 6) suggests that the reduced activity in response to vowels in children with ASD is not inherited from the earlier stages of analysis, such as tonotopic processing of formant frequencies [83], detection of harmonicity in complex sounds [84] or detection of an acoustic pattern [16]. On the other hand, timing of the vowel-related SPN reduction generally agrees with results of the meta-analysis of MMN/MMF studies which concluded that responses to phoneme changes (either vowels or syllables) are

reduced in individuals with ASD [32]. These considerations suggest that the processing deficit in children with ASD arises at the stage of phonetic analysis.

Notably, the time at which we observed a decrease in SPN in children with ASD coincides with the time at which categorization of isolated vowel sounds into distinct phoneme categories (e.g., /u/ vs. /a/) occurs (~ 175 ms post-stimulus onset) [85]. This stage is referred to as acoustic-phonetic mapping and distinguishes brain responses reflecting the true internalized percept of a vowel category from those that index acoustic properties of the vowel [85]. Deficits in phoneme category perception (e.g., relating vowels in the /y/ - /i/ continuum to the category /y/ or /i/) have previously been reported in children with ASD, despite preserved or even superior phoneme discrimination abilities (e.g., judging two vowels in the /y/ - /i/ continuum as the same or different) [86]. Therefore, it is likely that the neurofunctional abnormalities leading to decreased SPN in response to vowels in children with ASD reflect impaired acoustic-phonetic mapping necessary for phoneme categorization. This hypothesis is consistent with the observation that, in the left hemisphere, SPN reduction was most reliable and persistent in the mid-STG region located lateral to the primary auditory cortex in Heschl's gyrus (parabelt area A4 according to HCPMMP1 [57]) (Fig. 5). This area was suggested to be an initial STG waypoint of the ventral auditory stream – the auditory pathway optimized for recognition of acoustic pattern [13,87,88], especially those representing conspecific communication calls [89,90]. In humans, this parabelt auditory region plays a crucial role in phoneme encoding [13,91,92]. From a meta-analysis of neuroimaging studies of speech processing [13] concluded that phoneme recognition is associated with activation in the left mid-STG region, while the integration of phonemes into more complex patterns (i.e., words) is localized to the left anterior STG. This conclusion received strong support in a recent study of a patient with extensive lesions of the bilateral STS and left anterior STG, which showed that the intact region of the mid-STG alone can effectively subserve explicit vowel categorization despite the presence of "pure word deafness" [93].

The SPN in children with ASD was also decreased in the temporo-insular regions medially adjacent to the Hershel gyrus and extending in the anterior direction (Fig. 5). The pINS has strong structural and functional connections with the auditory cortex [59,94–97] and is responsive to a wide variety of acoustic stimuli [60]. Yet, registration of its neuronal responses in humans [98] and non-human primates [61] has shown that this auditory region of the pINS is sensitive to conspecific vocalizations and can transmit respective auditory information further down to the anterior insula, which is involved in the evaluation of affective signals conveyed by vocal sounds. In the future, it would be interesting to test whether reduced SPN in pINS regions is associated with impaired human voice emotion recognition in ASD [99].

Apart from signals of auditory modality, the pINS comprises neuronal representations of somatosensory, motor, visual, vestibular, limbic signals and is thought to be involved in multisensory integration [100]. The right insula seems to be particularly important for the audiovisual integration [101]. In this regard, it is interesting that the reduction in the SPN induced by vowels in our participants with ASD was strongest in the right pINS (Figure 5). Future studies should investigate whether atypical activity or connectivity of the right pINS contribute to severe deficit in audiovisual integration during phoneme recognition in children with ASD [102].

Suppressed processing of vowel formant structure is associated with words in noise perception difficulties in children with ASD

Importantly, the reduced negativity underlying processing of formant structure in children with ASD predicted the severity of their word recognition problems in the AM noise: the diminished SPN responses to vowels in the left hemisphere were associated with lower WiNam scores (Tab. 2, Fig.

7). This finding has several important implications for interpreting vowel processing deficit and its impact on auditory speech recognition in ASD.

First, while WiN performance in children with ASD showed some developmental improvement throughout childhood, neither child's age nor IQ could explain correlations between the reduced SPN and lowered WiNam scores (Table 2). The lack of correlation between WiN scores and IQ agrees well with the previous findings on the presence of speech-in-noise recognition difficulties even in high-functioning individuals with ASD [4]. On the other hand, our results suggest that these problems may be caused, at least in part, by a deficient vowel processing in the cortical networkc.

The passive presentation of auditory stimuli and the presence of the SPN deficit at already ~ 150 - 200 ms after sound onset - i.e., at the preattentive stage of processing - make the potential contribution of higher-order factors such as voluntary attention or motivation unlikely. Yet, involuntary orienting of attention to auditory stimuli may still influence differences between ASD and TD groups. Indeed, the P3a-like responses to periodic and nonperiodic vowels were observed in both TD and ASD children in our study, likely reflecting involuntary shift of attention to perceptually salient speech stimuli [103]. These responses were left-lateralized, consistent with the left-hemispheric bias of the P3a novelty response to speech revealed in the auditory cortex during intracranial recordings in patients [71]. The presence of elevated P3a to periodic vowels in the left hemisphere in our participants with ASD suggests that their reduced negative responses to vowels is unlikely to be due to inattention to the auditory stream containing speech sounds as it was previously suggested [104]. On the contrary, their involuntary attention seems to be captured by perceptually salient periodic vowel stimuli to a greater degree than in TD children.

The excessive P3a-like response could contribute to the decrease in the left-hemispheric SPN to periodic vowels and its correlation with WiNam scores in children with ASD, but it can hardly explain the general trend toward SPN reduction or a common correlation pattern for both periodic and non-periodic vowels. There are several arguments in support of this assumption. *(1)* No group differences in the P3a-like responses or distinct P3a-like peaks were observed in the right hemisphere, despite the prominent right-hemispheric SPN attenuation in ASD vs TD group (Fig. 4, 6). *(2)* In case of non-periodic vowels, the group differences in vowel-related negativity started already around 150 ms (Fig. 6D), i.e. in the time interval when P3a is not yet evident. *(3)* The group differences in P3a amplitude and in frequency-of-occurrence of the P3a peak were found for periodic vowels only, while in ASD group, the SPN was reduced for both periodic and nonperiodic vowels. *(4)* In the left hemisphere, SPN was a better predictor of WINam scores in children with ASD than the amplitude of the P3a-like component (Supplementary Table S2).

Although beyond the scope of this paper, the possible role of an enhanced left-hemispheric P3a-like response to speech sounds in autism deserves mention. The previous studies have shown that the P3a can be relatively independent of antecedent negativity. For example, Torppa et al. [105] observed in children with cochlear implants and normal hearing smaller MMN but larger P3a in response to speech sounds. Vlaskamp et al. [106] found that tone duration deviants induced smaller MMN, but larger P3a in children with ASD. It has been hypothesized that the larger P3a reflects increased recruitment of neural resources to compensate for less efficient automatic processing of salient sounds that lie outside the current attentional focus [106].

Second, despite the presence of altered vowel-evoked SPN in the auditory cortex of both hemispheres (Fig. 5), correlations with WiNam scores were only found in the left hemispheric (Table 2), indicating a specific relationship between the functional integrity of the left secondary auditory cortex and the ability to recognize words in fluctuating noise in children with ASD. Our previous study, which used the same stimuli to compare the SPN responses in neurotypical children and adults [25],

showed that the left hemispheric asymmetry in vowel-evoked SPN was present in adults but not in children, in whom SPN responses had the equal amplitude in both hemispheres. The correlation between left-hemispheric but not right-hemispheric SPN to vowels and WiNam scores in children with ASD suggests that some degree of left-hemispheric specialization for vowel processing is already present in childhood and possibly increases with age, driven by the need to integrate the encoding of vowel spectral composition with a predominantly left-lateralized language system (see [107] for the concept of graded hemispheric specialization).

The left STG region that most reliably separated between ASD and TD participants in the present study is remarkably similar to the region that appeared to be sensitive to intelligibility of the sentences that, in turn, depends on the slow temporal modulation of the speech signals at the level of syllables (3-4 Hz) [87]. Our results do not exclude a role of the left A4 region in sentence intelligibility, perhaps in the context of the top-down interactions between phonetic and higher-level (lexical, syntactic, working memory, etc.) processes, but suggest that this region is more specifically tuned to spectro-temporal composition of vowels and that weakening of this tuning hinders WiNam task performance.

Third, while the atypical left-hemispheric processing of vowels in children with ASD correlated with WiNam scores, it did not correlate with WiNst scores (Table 2). This pattern of correlations suggests that the impaired vowel processing interferes with the ability of listeners with ASD to use dips in noise to capture acoustic cues. Psychoacoustic studies have shown that in subjects with normal hearing, information important for word recognition in fluctuating noise is conveyed through both the temporal fine structure (TFS) of vowels, i.e., carrier frequencies of formants, and their common amplitude envelope (f0 / pitch) [36]. In our study, WiNam scores in children with ASD correlated with SPN evoked by periodic and non-periodic vowels (Table 2). Since both have a formant structure, but nonperiodic vowels lack *f0*, the atypical processing of formant frequencies seem to be the crucial factor contributing to the difficulties in perceiving word in AM noise in children with ASD. However, late (> 150 ms post-stimulus onset) occurrence of SPN reduction and its location in non-primary auditory cortex (left mid-STG region) suggests that the poor 'dip listening' is due to insufficient grouping of formants into a 'vowel object' rather than decoding of TFS separately for each formant frequency. Consistent with this hypothesis, a recent study of older adults found that impaired central grouping of acoustic patterns is a major contributor to their deficits in processing speech in noise [108].

Direction for future research

Word recognition in amplitude-modulated noise depends on multiple integrative processes occurring at different levels of the brain hierarchy and involving numerous feedforward, recurrent, and top-down interactions [109,110]. In a highly heterogeneous ASD population, difficulties with speech perception in noise may arise for a variety of reasons that are attributable to impairments at different stages of the auditory pathway or at higher hierarchical levels. Thus, our results indicating a role for impaired processing of vowel formant structure in WiN perception deficits in children with ASD do not exclude the contribution of other factors. In some children with autism, poor WiN recognition may be due to deficits occurring already at the subcortical level [111–115], as indexed by the frequency following response to speech sounds [116]. In the future, it would be important to investigate whether impairments in the analysis of the fine temporal structure of sound (TFS) [117] in the brainstem, and the deficit in cortical processes leading to the formation of auditory object contribute independently to poor WiN performance in children with ASD. Our findings do not rule out the "cognitive" hypothesis, which, based on behavioral results, attributes poor "masking release" (i.e., improved performance in AM compared to ST-noise) in individuals with ASD to a weakness of the

domain-unspecific mechanisms that integrate glimpsed fragments into meaningful speech [39,40]. However, since our study was not designed to test this hypothesis, additional neuroimaging research is warranted to address this issue.

Impaired speech-in-noise hearing is one of the central symptoms of auditory processing disorder - difficulties in recognizing and interpreting sounds, especially speech sounds that arise as a result of central nervous system dysfunction [118] and are frequently observed in children with ASD and other neurodevelopmental conditions [119]. Detecting at what CNS level the dysfunction takes place is important for development of effective and personalized intervention for auditory processing abnormalities not only in ASD, but also in other neurodevelopmental disorders. In this respect, our findings contribute to an emerging profile of children with developmental listening difficulties that suggests considerable variation in the neural causes underlying them.

LIMITATION

Our study has several limitations.

First, we restricted the analysis to temporal cortical regions, where the amplitude of response to sound is maximal, whereas important differences in the processing of linguistic stimuli in autism can also be observed outside the auditory cortex, such as in inferior frontal regions [120].

Second, we presented vowel stimuli which are very special overlearned conspecific auditory objects. It would be important to clarify whether the ASD-related deficit in SPN is specific to vowels or whether it is also observed for other auditory objects that have constant or coherently changing frequency composition.

Third, since we used simple words to test tor speech-in-noise processing difficulties in ASD, one should be cautious about generalizing the findings to more complex linguistic constructions such as sentences. In the case of sentences, speech recognition in noise may be supported by prosody, the slow (syllable-rate) envelope of the speech signal [36], and higher-order semantic cues [121] that are absent or less important in the case of isolated words.

Fourth, we did not control subjects' attention to the auditory stimuli, so the possibility remains that differences in attention allocation could affect the results. Comparing the responses in passive and active listening paradigms may help to clarify the role of attention in the observed differences in SPN.

CONCLUSION

Our results suggest that a substantial proportion of children with ASD have altered functional integrity of non-primary auditory cortical areas involved in processing of the vowel formant composition. The localization and relatively late occurrence of this deficit suggest that it arises at the stage of integration of individual formants into phonetic objects – vowels – whereas no deficit was found in children with ASD at earlier stages of processing associated with the encoding of individual formants and/or the detection of a frequency pattern. In the left hemisphere, neural deficit in vowel processing was associated with difficulty recovering words from glimpses of speech in fluctuating noise, a problem characteristic of children with autism. Thus, along with the low- and high-level speech processing impairments hypothesized to be characteristic of ASD individuals, the impaired grouping of acoustic features into phonetic objects has an adverse effect on the ability to recognise speech in fluctuating noise in children with autism.

DECLARATIONS

Acknowledgments

We sincerely thank children and their families for participating in this study. The study was conducted at the unique research facility “Center for Neurocognitive Research (MEG-Center)” of MSUPE.

Funding

The study was funded within the framework of the state assignment of the Ministry of Education of the Russian Federation (№ 073-00037-24-01).

Authors' contributions

Fadeev K.A.: Formal analysis, Data recording, Data curation, Investigation, Visualization, Software, Writing - original draft; Writing - review & editing;
Romero Reyes, I.V.: Formal analysis, Data curation, Investigation, Visualization, Software, Writing - original draft; Writing - review & editing;
Goiaeva D.E.: Data recording, Investigation, Project administration, Writing - review & editing;
Obukhova T.S.: Data recording, Investigation, Project administration, Writing - review & editing;
Ovsiannikova T.M.: Data recording, Investigation, Project administration, Writing - review & editing;
Prokofyev A.O.: Data recording, Investigation, Writing - review & editing;
Rytikova A.M.: Investigation, Writing - review & editing;
Novikov A.Y.: Investigation, Writing - review & editing;
Kozunov V.V.: Investigation, Writing - review & editing;
Stroganova T.A.: Conceptualization, Methodology, Funding acquisition, Writing - original draft; Writing - review & editing.
Orekhova E.V.: Conceptualization, Methodology, Formal analysis, Visualization, Funding acquisition, Writing - original draft, Writing - review & editing;

Consent for publication

All children gave verbal consent to participate in the study and their caregivers gave written consent for publication of anonymized data.

Competing interests

The authors declare that they have no competing interests.

Data/code availability statement

The de-identified individual-level raw data that supports this research, study materials, and analysis code used to generate the results are publicly available at https://openneuro.org/datasets/ds005234

REFERENCES

1. American Psychiatric Association. Diagnostic and statistical manual of mental disorders (5th ed.) [Internet]. 2013. Available from: http://dx.doi.org/10.1176/appi.books.9780890425596

2. Tager-Flusberg H, Kasari C. Minimally verbal school-aged children with autism spectrum disorder: the neglected end of the spectrum. Autism Res [Internet]. 2013;6(6):468–78. Available from: http://dx.doi.org/10.1002/aur.1329

3. Mody M, Belliveau JW. Speech and Language Impairments in Autism: Insights from Behavior and Neuroimaging. N Am J Med Sci [Internet]. 2013;5(3):157–61. Available from:

http://dx.doi.org/10.7156/v5i3p157

4. Ruiz Callejo D, Boets B. A systematic review on speech-in-noise perception in autism. Neurosci Biobehav Rev [Internet]. 2023;154:105406. Available from: http://dx.doi.org/10.1016/j.neubiorev.2023.105406

5. Sturrock A, Guest H, Hanks G, Bendo G, Plack CJ, Gowen E. Chasing the conversation: Autistic experiences of speech perception. Autism & developmental language impairments [Internet]. 2022;7:23969415221077532. Available from: http://dx.doi.org/10.1177/23969415221077532

6. Eigsti IM, de Marchena AB, Schuh JM, Kelley E. Language Acquisition in Autism Spectrum Disorders: A Developmental Review. Res Autism Spectr Disord [Internet]. 2011;5:681–91. Available from: http://dx.doi.org/10.1016/j.rasd.2010.09.001

7. Key AP, D'Ambrose Slaboch K. Speech Processing in Autism Spectrum Disorder: An Integrative Review of Auditory Neurophysiology Findings. J Speech Lang Hear Res [Internet]. 2021;64(11):4192–212. Available from: http://dx.doi.org/10.1044/2021_jslhr-20-00738

8. Shokrkon A, Nicoladis E. The Directionality of the Relationship Between Executive Functions and Language Skills: A Literature Review. Front Psychol [Internet]. 2022;13:848696. Available from: http://dx.doi.org/10.3389/fpsyg.2022.848696

9. Kent RD, Rountrey C. What Acoustic Studies Tell Us About Vowels in Developing and Disordered Speech. Am J Speech Lang Pathol [Internet]. 2020;29(3):1749–78. Available from: http://dx.doi.org/10.1044/2020_ajslp-19-00178

10. Boë LJ, Berthommier F, Legou T, Captier G, Kemp C, Sawallis TR, et al. Evidence of a Vocalic Proto-System in the Baboon (Papio papio) Suggests Pre-Hominin Speech Precursors. PLoS One [Internet]. 2017;12(1):e0169321. Available from: http://dx.doi.org/10.1371/journal.pone.0169321

11. Bizley JK, Cohen YE. The what, where and how of auditory-object perception. Nat Rev Neurosci [Internet]. 2013;14(10):693–707. Available from: http://dx.doi.org/10.1038/nrn3565

12. Brefczynski-Lewis JA, Lewis JW. Auditory object perception: A neurobiological model and prospective review. Neuropsychologia [Internet]. 2017;105:223–42. Available from: http://dx.doi.org/10.1016/j.neuropsychologia.2017.04.034

13. DeWitt I, Rauschecker JP. Phoneme and word recognition in the auditory ventral stream. Proc Natl Acad Sci U S A [Internet]. 2012;109(8):E505–14. Available from: http://dx.doi.org/10.1073/pnas.1113427109

14. Wang X, Lu T, Snider RK, Liang L. Sustained firing in auditory cortex evoked by preferred stimuli. Nature [Internet]. 2005;435(7040):341–6. Available from: http://dx.doi.org/10.1038/nature03565

15. Wang X. Cortical Coding of Auditory Features. Annu Rev Neurosci [Internet]. 2018;41:527–52. Available from: http://dx.doi.org/10.1146/annurev-neuro-072116-031302

16. Molloy K, Lavie N, Chait M. Auditory Figure-Ground Segregation Is Impaired by High Visual Load. J Neurosci [Internet]. 2019;39(9):1699–708. Available from: http://dx.doi.org/10.1523/jneurosci.2518-18.2018

17. Chait M. How the brain discovers structure in sound sequences. Acoust Sci Technol [Internet]. 2020;41(1):48–53. Available from: http://dx.doi.org/10.1250/ast.41.48

18. van Dommelen WA. Acoustic parameters in human speaker recognition. Lang Speech [Internet]. 1990;33 (Pt 3):259–72. Available from: http://dx.doi.org/10.1177/002383099003300302

19. Lin IF, Yamada T, Komine Y, Kato N, Kashino M. Enhanced segregation of concurrent sounds with similar spectral uncertainties in individuals with autism spectrum disorder. Sci Rep [Internet]. 2015;5:10524. Available from: http://dx.doi.org/10.1038/srep10524

20. Lin IF, Shirama A, Kato N, Kashino M. The singular nature of auditory and visual scene analysis in autism. Philos Trans R Soc Lond B Biol Sci [Internet]. 2017;372(1714). Available from: http://dx.doi.org/10.1098/rstb.2016.0115

21. Bharadwaj H, Mamashli F, Khan S, Singh R, Joseph RM, Losh A, et al. Cortical signatures of auditory object binding in children with autism spectrum disorder are anomalous in concordance with behavior and diagnosis. PLoS Biol [Internet]. 2022;20(2):e3001541. Available from: http://dx.doi.org/10.1371/journal.pbio.3001541

22. Barascud N, Pearce MT, Griffiths TD, Friston KJ, Chait M. Brain responses in humans reveal ideal observer-like sensitivity to complex acoustic patterns. Proc Natl Acad Sci U S A [Internet]. 2016;113(5):E616–25. Available from: http://dx.doi.org/10.1073/pnas.1508523113

23. Gutschalk A, Uppenkamp S. Sustained responses for pitch and vowels map to similar sites in human auditory cortex. Neuroimage [Internet]. 2011;56(3):1578–87. Available from: http://dx.doi.org/10.1016/j.neuroimage.2011.02.026

24. Teki S, Barascud N, Picard S, Payne C, Griffiths TD, Chait M. Neural Correlates of Auditory Figure-Ground Segregation Based on Temporal Coherence. Cereb Cortex [Internet]. 2016;26(9):3669–80. Available from: http://dx.doi.org/10.1093/cercor/bhw173

25. Orekhova EV, Fadeev KA, Goiaeva DE, Obukhova TS, Ovsiannikova TM, Prokofyev AO, et al. Different hemispheric lateralization for periodicity and formant structure of vowels in the auditory cortex and its changes between childhood and adulthood. Cortex [Internet]. 2024;171:287–307. Available from: http://dx.doi.org/10.1016/j.cortex.2023.10.020

26. Steinmann I, Gutschalk A. Sustained BOLD and theta activity in auditory cortex are related to slow stimulus fluctuations rather than to pitch. J Neurophysiol [Internet]. 2012;107(12):3458–67. Available from: http://dx.doi.org/10.1152/jn.01105.2011

27. Tian B, Rauschecker JP. Processing of frequency-modulated sounds in the lateral auditory belt cortex of the rhesus monkey. J Neurophysiol [Internet]. 2004;92(5):2993–3013. Available from: http://dx.doi.org/10.1152/jn.00472.2003

28. Walker KM, Schnupp JW, Hart-Schnupp SM, King AJ, Bizley JK. Pitch discrimination by ferrets for simple and complex sounds. J Acoust Soc Am [Internet]. 2009;126(3):1321–35. Available from: http://dx.doi.org/10.1121/1.3179676

29. Walker KM, Bizley JK, King AJ, Schnupp JW. Cortical encoding of pitch: recent results and open questions. Hear Res [Internet]. 2011;271(1-2):74–87. Available from: http://dx.doi.org/10.1016/j.heares.2010.04.015

30. Sadagopan S, Wang X. Level invariant representation of sounds by populations of neurons in primary auditory cortex. J Neurosci [Internet]. 2008;28(13):3415–26. Available from: http://dx.doi.org/10.1523/jneurosci.2743-07.2008

31. Stroganova TA, Komarov KS, Goyaeva DE, Obukhova TS, Ovsyannikova TM, Prokofiev AO, et al. The effect of periodicity and “vowelness” of a sound on cortical auditory responses in children. IP Pavlov Journal of Higher Nervous Activity [Internet]. 2021;71(4). Available from: http://dx.doi.org/10.31857/S00

32. Chen TC, Hsieh MH, Lin YT, Chan PS, Cheng CH. Mismatch negativity to different deviant changes in autism spectrum disorders: A meta-analysis. Clin Neurophysiol [Internet].

2020;131(3):766–77. Available from: http://dx.doi.org/10.1016/j.clinph.2019.10.031

33. Ponton C, Eggermont JJ, Khosla D, Kwong B, Don M. Maturation of human central auditory system activity: separating auditory evoked potentials by dipole source modeling. Clin Neurophysiol [Internet]. 2002;113(3):407–20. Available from: http://dx.doi.org/10.1016/s1388-2457(01)00733-7

34. Parviainen T, Helenius P, Salmelin R. Children show hemispheric differences in the basic auditory response properties. Hum Brain Mapp [Internet]. 2019;40(9):2699–710. Available from: http://dx.doi.org/10.1002/hbm.24553

35. Orekhova EV, Butorina AV, Tsetlin MM, Novikova SI, Sokolov PA, Elam M, et al. Auditory magnetic response to clicks in children and adults: its components, hemispheric lateralization and repetition suppression effect. Brain Topogr [Internet]. 2013;26(3):410–27. Available from: http://dx.doi.org/10.1007/s10548-012-0262-x

36. Fogerty D, Humes LE. The role of vowel and consonant fundamental frequency, envelope, and temporal fine structure cues to the intelligibility of words and sentences. J Acoust Soc Am [Internet]. 2012;131(2):1490–501. Available from: http://dx.doi.org/10.1121/1.3676696

37. Fadeev KA, Goyaeva DE, Obukhova TS, Ovsyannikova TM, Shvedovskiy EF, Yu. Nikolaeva A, et al. Difficulty with Speech Perception in the Background of Noise in Children with Autism Spectrum Disorders Is Not Related to Their Level of Intelligence. Clinical Psychology and Special Education [Internet]. 2023 [cited 2024 May 24];12(1):180–212. Available from: https://psyjournals.ru/journals/cpse/archive/2023_n1/cpse_2023_n1_Fadeev_et_al.pdf

38. Hopkins K, Moore BC. The contribution of temporal fine structure to the intelligibility of speech in steady and modulated noise. J Acoust Soc Am [Internet]. 2009;125(1):442–6. Available from: http://dx.doi.org/10.1121/1.3037233

39. Groen WB, van Orsouw L, Huurne N, Swinkels S, van der Gaag RJ, Buitelaar JK, et al. Intact spectral but abnormal temporal processing of auditory stimuli in autism. J Autism Dev Disord [Internet]. 2009;39(5):742–50. Available from: http://dx.doi.org/10.1007/s10803-008-0682-3

40. Alcántara JI, Weisblatt EJ, Moore BC, Bolton PF. Speech-in-noise perception in high-functioning individuals with autism or Asperger's syndrome. J Child Psychol Psychiatry [Internet]. 2004;45(6):1107–14. Available from: http://dx.doi.org/10.1111/j.1469-7610.2004.t01-1-00303.x

41. Constantino JN. Social Responsiveness Scale. In: Volkmar FR, editor. Encyclopedia of Autism Spectrum Disorders [Internet]. New York, NY: Springer New York; 2013. p. 2919–29. Available from: https://doi.org/10.1007/978-1-4419-1698-3_296

42. Berument SK, Rutter M, Lord C, Pickles A, Bailey A. Autism screening questionnaire: diagnostic validity. Br J Psychiatry [Internet]. 1999;175:444–51. Available from: http://dx.doi.org/10.1192/bjp.175.5.444

43. Kaufman AS, Kaufman NL. KABC-II : Kaufman Assessment Battery for Children. 2nd ed. Circle Pines, MN: AGS Pub.; 2004.

44. Drozdick LW, Singer JK, Lichtenberger EO, Kaufman JC, Kaufman AS, Kaufman NL. The Kaufman Assessment Battery for Children—Second Edition and KABC-II Normative Update. In: D. P. F, McDonough EM, editors. Contemporary intellectual assessment: Theories, tests, and issues. The Guilford Press; 2018. p. 333–59.

45. British Society of Audiology. Recommended procedure: pure-tone air-conduction and boneconduction threshold audiometry with and without masking. 2018; Available from: https://www.thebsa.org.uk/wp-content/uploads/2023/10/OD104-32-Recommended-Procedure-

Pure-Tone-Audiometry-August-2018-FINAL-1.pdf

46. Lopukhina A, Chrabaszcz A, Khudyakova M, Korkina I, Yurchenko A, Dragoy O. Test for assessment of language development in Russian «KORABLIK». 2019. 30 p.

47. Ivanova, M., Dragoy, O., Akinina, J., et al. AutoRAT as your fingertips: Introducing the new Russian Aphasia Test on tablet. Front Psychol [Internet]. 2016;2016(1). Available from: http://dx.doi.org/10.3389/conf.fpsyg.2016.68.00116

48. Lyashevskaya O. N. SSA. Chastotnyi slovar' sovremennogo russkogo yazyka (na materialakh Natsional'nogo korpusa russkogo yazyka) [Frequency Dictionary of the Modern Russian Language (based on the materials of the National Corpus of the Russian Language)] [Internet]. Moscow: Azbukovnik; 2009. Available from: https://elibrary.ru/item.asp?id=21067600

49. Partington JW, Sundberg ML. The Assessment of Basic Language and Learning Skills. Pleasant Hill, CA: Behavior Analysts, Inc.; 1998.

50. Greene C. Sound Pressure Level Calculator [Internet]. MATLAB Central File Exchange; 2016. Available from: https://www.mathworks.com/matlabcentral/fileexchange/35876-sound-pressurelevel-calculator

51. Uppenkamp S, Johnsrude IS, Norris D, Marslen-Wilson W, Patterson RD. Locating the initial stages of speech-sound processing in human temporal cortex. Neuroimage [Internet]. 2006 Jul 1;31(3):1284–96. Available from: http://dx.doi.org/10.1016/j.neuroimage.2006.01.004

52. Dale AM, Fischl B, Sereno MI. Cortical Surface-Based Analysis: I. Segmentation and Surface Reconstruction. Neuroimage [Internet]. 1999 Feb 1;9(2):179–94. Available from: https://www.sciencedirect.com/science/article/pii/S1053811998903950

53. Fischl B, Sereno MI, Dale AM. Cortical Surface-Based Analysis: II: Inflation, Flattening, and a Surface-Based Coordinate System. Neuroimage [Internet]. 1999 Feb 1;9(2):195–207. Available from: https://www.sciencedirect.com/science/article/pii/S1053811998903962

54. Larson E, Gramfort A, Engemann DA, Leppakangas J, Brodbeck C, Jas M, et al. MNE-Python [Internet]. Available from: https://zenodo.org/records/8262486

55. Pascual-Marqui RD. Standardized low-resolution brain electromagnetic tomography (sLORETA): technical details. Methods Find Exp Clin Pharmacol [Internet]. 2002;24 Suppl D:5–12. Available from: https://www.ncbi.nlm.nih.gov/pubmed/12575463

56. Smith SM, Nichols TE. Threshold-free cluster enhancement: addressing problems of smoothing, threshold dependence and localisation in cluster inference. Neuroimage [Internet]. 2009;44(1):83–98. Available from: http://dx.doi.org/10.1016/j.neuroimage.2008.03.061

57. Glasser MF, Coalson TS, Robinson EC, Hacker CD, Harwell J, Yacoub E, et al. A multi-modal parcellation of human cerebral cortex. Nature [Internet]. 2016;536(7615):171–8. Available from: http://dx.doi.org/10.1038/nature18933

58. Baker CM, Burks JD, Briggs RG, Conner AK, Glenn CA, Robbins JM, et al. A Connectomic Atlas of the Human Cerebrum-Chapter 5: The Insula and Opercular Cortex. Operative neurosurgery (Hagerstown, Md) [Internet]. 2018;15(suppl_1):S175–s244. Available from: http://dx.doi.org/10.1093/ons/opy259

59. Rolls ET, Rauschecker JP, Deco G, Huang CC, Feng J. Auditory cortical connectivity in humans. Cereb Cortex [Internet]. 2023 May 9;33(10):6207–27. Available from: http://dx.doi.org/10.1093/cercor/bhac496

60. Blenkmann AO, Collavini S, Lubell J, Llorens A, Funderud I, Ivanovic J, et al. Auditory deviance detection in the human insula: An intracranial EEG study. Cortex [Internet]. 2019;121:189–200. Available from: http://dx.doi.org/10.1016/j.cortex.2019.09.002

61. Remedios R, Logothetis NK, Kayser C. An auditory region in the primate insular cortex responding preferentially to vocal communication sounds. J Neurosci [Internet]. 2009;29(4):1034–45. Available from: http://dx.doi.org/10.1523/jneurosci.4089-08.2009

62. Tamura Y, Kuriki S, Nakano T. Involvement of the left insula in the ecological validity of the human voice. Sci Rep [Internet]. 2015;5:8799. Available from: http://dx.doi.org/10.1038/srep08799

63. Hauk O, Stenroos M, Treder MS. Towards an objective evaluation of EEG/MEG source estimation methods - The linear approach. Neuroimage [Internet]. 2022;255:119177. Available from: http://dx.doi.org/10.1016/j.neuroimage.2022.119177

64. Das D, Shaw ME, Hämäläinen MS, Dykstra AR, Doll L, Gutschalk A. A role for retro-splenial cortex in the task-related P3 network. Clin Neurophysiol [Internet]. 2024;157:96–109. Available from: http://dx.doi.org/10.1016/j.clinph.2023.11.014

65. Alho J, Bharadwaj H, Khan S, Mamashli F, Perrachione TK, Losh A, et al. Altered maturation and atypical cortical processing of spoken sentences in autism spectrum disorder. Prog Neurobiol [Internet]. 2021;203:102077. Available from: http://dx.doi.org/10.1016/j.pneurobio.2021.102077

66. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. 1995;57:289–300.

67. Aldridge FJ, Gibbs VM, Schmidhofer K, Williams M. Investigating the clinical usefulness of the Social Responsiveness Scale (SRS) in a tertiary level, autism spectrum disorder specific assessment clinic. J Autism Dev Disord [Internet]. 2012;42(2):294–300. Available from: http://dx.doi.org/10.1007/s10803-011-1242-9

68. Corsello C, Hus V, Pickles A, Risi S, Cook EH Jr, Leventhal BL, et al. Between a ROC and a hard place: decision making and making decisions about using the SCQ. J Child Psychol Psychiatry [Internet]. 2007;48(9):932–40. Available from: http://dx.doi.org/10.1111/j.1469-7610.2007.01762.x

69. Hämäläinen M, Hari R. 10 - Magnetoencephalographic Characterization of Dynamic Brain Activation: Basic Principles and Methods of Data Collection and Source Analysis. In: Toga AW, Mazziotta JC, editors. Brain Mapping: The Methods (Second Edition) [Internet]. San Diego: Academic Press; 2002. p. 227–53. Available from: https://www.sciencedirect.com/science/article/pii/B9780126930191500125

70. Hewson-Stoate N, Schönwiesner M, Krumbholz K. Vowel processing evokes a large sustained response anterior to primary auditory cortex. Eur J Neurosci [Internet]. 2006;24(9):2661–71. Available from: http://dx.doi.org/10.1111/j.1460-9568.2006.05096.x

71. Nourski KV, Steinschneider M, Rhone AE, Dappen ER, Kawasaki H, Howard MA 3rd. Processing of auditory novelty in human cortex during a semantic categorization task. Hear Res [Internet]. 2024;444:108972. Available from: http://dx.doi.org/10.1016/j.heares.2024.108972

72. Lenhard W, Lenhard A. Hypothesis Tests for Comparing Correlations [Internet]. 2014 [cited 2024 May 25]. Available from: http://www.psychometrica.de/correlation.html

73. Keceli S, Okamoto H, Kakigi R. Hierarchical Neural Encoding of Temporal Regularity in the Human Auditory Cortex. Brain Topogr [Internet]. 2015;28(3):459–70. Available from: http://dx.doi.org/10.1007/s10548-013-0300-3

74. Southwell R, Chait M. Enhanced deviant responses in patterned relative to random sound sequences. Cortex [Internet]. 2018 Dec;109:92–103. Available from: http://dx.doi.org/10.1016/j.cortex.2018.08.032

75. Hu M, Bianco R, Hidalgo AR, Chait M. Concurrent Encoding of Sequence Predictability and Event-Evoked Prediction Error in Unfolding Auditory Patterns. J Neurosci [Internet]. 2024;44(14):e1894232024. Available from: http://dx.doi.org/10.1523/jneurosci.1894-23.2024

76. Tóth B, Kocsis Z, Háden GP, Szerafin Á, Shinn-Cunningham BG, Winkler I. EEG signatures accompanying auditory figure-ground segregation. Neuroimage [Internet]. 2016;141:108–19. Available from: http://dx.doi.org/10.1016/j.neuroimage.2016.07.028

77. Belin P, Fecteau S, Bédard C. Thinking the voice: neural correlates of voice perception. Trends Cogn Sci [Internet]. 2004;8(3):129–35. Available from: http://dx.doi.org/10.1016/j.tics.2004.01.008

78. Bethmann A, Brechmann A. On the definition and interpretation of voice selective activation in the temporal cortex. Front Hum Neurosci [Internet]. 2014;8:499. Available from: http://dx.doi.org/10.3389/fnhum.2014.00499

79. Pernet CR, McAleer P, Latinus M, Gorgolewski KJ, Charest I, Bestelmeyer PE, et al. The human voice areas: Spatial organization and inter-individual variability in temporal and extra-temporal cortices. Neuroimage [Internet]. 2015;119:164–74. Available from: http://dx.doi.org/10.1016/j.neuroimage.2015.06.050

80. Samson F, Hyde KL, Bertone A, Soulières I, Mendrek A, Ahad P, et al. Atypical processing of auditory temporal complexity in autistics. Neuropsychologia [Internet]. 2011;49(3):546–55. Available from: http://dx.doi.org/10.1016/j.neuropsychologia.2010.12.033

81. Lai G, Schneider HD, Schwarzenberger JC, Hirsch J. Speech stimulation during functional MR imaging as a potential indicator of autism. Radiology [Internet]. 2011;260(2):521–30. Available from: http://dx.doi.org/10.1148/radiol.11101576

82. Engineer CT, Centanni TM, Im KW, Borland MS, Moreno NA, Carraway RS, et al. Degraded auditory processing in a rat model of autism limits the speech representation in non-primary auditory cortex. Dev Neurobiol [Internet]. 2014 Oct;74(10):972–86. Available from: http://dx.doi.org/10.1002/dneu.22175

83. Fisher JM, Dick FK, Levy DF, Wilson SM. Neural representation of vowel formants in tonotopic auditory cortex. Neuroimage [Internet]. 2018;178:574–82. Available from: http://dx.doi.org/10.1016/j.neuroimage.2018.05.072

84. Feng L, Wang X. Harmonic template neurons in primate auditory cortex underlying complex sound processing. Proc Natl Acad Sci U S A [Internet]. 2017;114(5):E840–e848. Available from: http://dx.doi.org/10.1073/pnas.1607519114

85. Bidelman GM, Moreno S, Alain C. Tracing the emergence of categorical speech perception in the human auditory system. Neuroimage [Internet]. 2013 Oct 1;79:201–12. Available from: http://dx.doi.org/10.1016/j.neuroimage.2013.04.093

86. You RS, Serniclaes W, Rider D, Chabane N. On the nature of the speech perception deficits in children with autism spectrum disorders. Res Dev Disabil [Internet]. 2017 Feb;61:158–71. Available from: http://dx.doi.org/10.1016/j.ridd.2016.12.009

87. Albouy P, Benjamin L, Morillon B, Zatorre RJ. Distinct sensitivity to spectrotemporal modulation supports brain asymmetry for speech and melody. Science [Internet]. 2020 Feb 28;367(6481):1043–7. Available from: http://dx.doi.org/10.1126/science.aaz3468

88. Romanski LM, Averbeck BB. The primate cortical auditory system and neural representation of conspecific vocalizations. Annu Rev Neurosci [Internet]. 2009;32:315–46. Available from: http://dx.doi.org/10.1146/annurev.neuro.051508.135431

89. Rauschecker JP, Tian B, Hauser M. Processing of complex sounds in the macaque nonprimary auditory cortex. Science [Internet]. 1995;268(5207):111–4. Available from: http://dx.doi.org/10.1126/science.7701330

90. Rauschecker JP. Ventral and dorsal streams in the evolution of speech and language. Front Evol Neurosci [Internet]. 2012;4:7. Available from: http://dx.doi.org/10.3389/fnevo.2012.00007

91. Norman-Haignere S, Kanwisher NG, McDermott JH. Distinct Cortical Pathways for Music and Speech Revealed by Hypothesis-Free Voxel Decomposition. Neuron [Internet]. 2015;88(6):1281–96. Available from: http://dx.doi.org/10.1016/j.neuron.2015.11.035

92. Khalighinejad B, Patel P, Herrero JL, Bickel S, Mehta AD, Mesgarani N. Functional characterization of human Heschl's gyrus in response to natural speech. Neuroimage [Internet]. 2021 Jul 15;235:118003. Available from: http://dx.doi.org/10.1016/j.neuroimage.2021.118003

93. Gutschalk A, Uppenkamp S, Riedel B, Bartsch A, Brandt T, Vogt-Schaden M. Pure word deafness with auditory object agnosia after bilateral lesion of the superior temporal sulcus. Cortex [Internet]. 2015 Dec;73:24–35. Available from: http://dx.doi.org/10.1016/j.cortex.2015.08.001

94. Dionisio S, Mayoglou L, Cho SM, Prime D, Flanigan PM, Lega B, et al. Connectivity of the human insula: A cortico-cortical evoked potential (CCEP) study. Cortex [Internet]. 2019;120:419–42. Available from: http://dx.doi.org/10.1016/j.cortex.2019.05.019

95. Zachlod D, Rüttgers B, Bludau S, Mohlberg H, Langner R, Zilles K, et al. Four new cytoarchitectonic areas surrounding the primary and early auditory cortex in human brains. Cortex [Internet]. 2020;128:1–21. Available from: http://dx.doi.org/10.1016/j.cortex.2020.02.021

96. Ghaziri J, Tucholka A, Girard G, Houde JC, Boucher O, Gilbert G, et al. The Corticocortical Structural Connectivity of the Human Insula. Cereb Cortex [Internet]. 2017;27(2):1216–28. Available from: http://dx.doi.org/10.1093/cercor/bhv308

97. Uddin LQ, Nomi JS, Hébert-Seropian B, Ghaziri J, Boucher O. Structure and Function of the Human Insula. J Clin Neurophysiol [Internet]. 2017;34(4):300–6. Available from: http://dx.doi.org/10.1097/wnp.0000000000000377

98. Zhang Y, Zhou W, Huang J, Hong B, Wang X. Neural correlates of perceived emotions in human insula and amygdala for auditory emotion recognition. Neuroimage [Internet]. 2022 Oct 15;260:119502. Available from: http://dx.doi.org/10.1016/j.neuroimage.2022.119502

99. Schelinski S, von Kriegstein K. The Relation Between Vocal Pitch and Vocal Emotion Recognition Abilities in People with Autism Spectrum Disorder and Typical Development. J Autism Dev Disord [Internet]. 2019;49(1):68–82. Available from: http://dx.doi.org/10.1007/s10803-018-3681-z

100. Bamiou DE, Musiek FE, Luxon LM. The insula (Island of Reil) and its role in auditory processing. Literature review. Brain Res Brain Res Rev [Internet]. 2003;42(2):143–54. Available from: http://dx.doi.org/10.1016/s0165-0173(03)00172-3

101. Bushara KO, Grafman J, Hallett M. Neural correlates of auditory-visual stimulus onset asynchrony detection. J Neurosci [Internet]. 2001 Jan 1;21(1):300–4. Available from: http://dx.doi.org/10.1523/JNEUROSCI.21-01-00300.2001

102. Foxe JJ, Molholm S, Del Bene VA, Frey HP, Russo NN, Blanco D, et al. Severe multisensory

speech integration deficits in high-functioning school-aged children with Autism Spectrum Disorder (ASD) and their resolution during early adolescence. Cereb Cortex [Internet]. 2015 Feb;25(2):298–312. Available from: http://dx.doi.org/10.1093/cercor/bht213

103. Escera C, Corral MJ. Role of Mismatch Negativity and Novelty-P3 in Involuntary Auditory Attention. J Psychophysiol [Internet]. 2007 Jan 1;21(3-4):251–64. Available from: https://doi.org/10.1027/0269-8803.21.34.251

104. Ceponiene R, Lepistö T, Shestakova A, Vanhala R, Alku P, Näätänen R, et al. Speech-sound-selective auditory impairment in children with autism: they can perceive but do not attend. Proc Natl Acad Sci U S A [Internet]. 2003;100(9):5567–72. Available from: http://dx.doi.org/10.1073/pnas.0835631100

105. Torppa R, Kuuluvainen S, Lipsanen J. The development of cortical processing of speech differs between children with cochlear implants and normal hearing and changes with parental singing. Front Neurosci [Internet]. 2022;16:976767. Available from: http://dx.doi.org/10.3389/fnins.2022.976767

106. Vlaskamp C, Oranje B, Madsen GF, Møllegaard Jepsen JR, Durston S, Cantio C, et al. Auditory processing in autism spectrum disorder: Mismatch negativity deficits. Autism Res [Internet]. 2017;10(11):1857–65. Available from: http://dx.doi.org/10.1002/aur.1821

107. Behrmann M, Plaut DC. A vision of graded hemispheric specialization. Ann N Y Acad Sci [Internet]. 2015;1359:30–46. Available from: http://dx.doi.org/10.1111/nyas.12833

108. Holmes E, Griffiths TD. “Normal” hearing thresholds and fundamental auditory grouping processes predict difficulties with speech-in-noise perception. Sci Rep [Internet]. 2019;9(1):16771. Available from: http://dx.doi.org/10.1038/s41598-019-53353-5

109. Felix RA 2nd, Gourévitch B, Portfors CV. Subcortical pathways: Towards a better understanding of auditory disorders. Hear Res [Internet]. 2018;362:48–60. Available from: http://dx.doi.org/10.1016/j.heares.2018.01.008

110. Rauschecker JP. Dual stream models of auditory vocal communication. In: The Oxford handbook of voice perception. Oxford University Press; 2018. p. 413.

111. Otto-Meyer S, Krizman J, White-Schwoch T, Kraus N. Children with autism spectrum disorder have unstable neural responses to sound. Exp Brain Res [Internet]. 2018;236(3):733–43. Available from: http://dx.doi.org/10.1007/s00221-017-5164-4

112. Lau JCY, To CKS, Kwan JSK, Kang X, Losh M, Wong PCM. Lifelong Tone Language Experience does not Eliminate Deficits in Neural Encoding of Pitch in Autism Spectrum Disorder. J Autism Dev Disord [Internet]. 2021;51(9):3291–310. Available from: http://dx.doi.org/10.1007/s10803-020-04796-7

113. Ramezani M, Lotfi Y, Moossavi A, Bakhshi E. Auditory brainstem response to speech in children with high functional autism spectrum disorder. Neurol Sci [Internet]. 2019;40(1):121–5. Available from: http://dx.doi.org/10.1007/s10072-018-3594-9

114. Chen J, Liang C, Wei Z, Cui Z, Kong X, Dong CJ, et al. Atypical longitudinal development of speech-evoked auditory brainstem response in preschool children with autism spectrum disorders. Autism Res [Internet]. 2019;12(7):1022–31. Available from: http://dx.doi.org/10.1002/aur.2110

115. Tecoulesco L, Skoe E, Naigles LR. Phonetic discrimination mediates the relationship between auditory brainstem response stability and syntactic performance. Brain Lang [Internet]. 2020;208:104810. Available from: http://dx.doi.org/10.1016/j.bandl.2020.104810

116. Krizman J, Kraus N. Analyzing the FFR: A tutorial for decoding the richness of auditory function. Hear Res [Internet]. 2019;382:107779. Available from: http://dx.doi.org/10.1016/j.heares.2019.107779

117. Moon IJ, Hong SH. What is temporal fine structure and why is it important? Korean journal of audiology [Internet]. 2014;18(1):1–7. Available from: http://dx.doi.org/10.7874/kja.2014.18.1.1

118. Aristidou IL, Hohman MH. Central Auditory Processing Disorder. StatPearls, Treasure Island (FL): StatPearls Publishing; 2023.

119. de Wit E, van Dijk P, Hanekamp S, Visser-Bochane MI, Steenbergen B, van der Schans CP, et al. Same or Different: The Overlap Between Children With Auditory Processing Disorders and Children With Other Developmental Disorders: A Systematic Review. Ear Hear [Internet]. 2018;39(1):1–19. Available from: http://dx.doi.org/10.1097/aud.0000000000000479

120. Yoshimura Y, Kikuchi M, Hayashi N, Hiraishi H, Hasegawa C, Takahashi T, et al. Altered human voice processing in the frontal cortex and a developmental language delay in 3- to 5-year-old children with autism spectrum disorder. Sci Rep [Internet]. 2017;7(1):17116. Available from: http://dx.doi.org/10.1038/s41598-017-17058-x

121. Shi LF. Measuring effectiveness of semantic cues in degraded English sentences in non-native listeners. Int J Audiol [Internet]. 2014;53(1):30–9. Available from: http://dx.doi.org/10.3109/14992027.2013.825052