

RESEARCH ARTICLE

Open Access



Spectral-temporal EEG dynamics of speech discrimination processing in infants during sleep

Phillip M. Gilley^{1,2*} , Kristin Uhler³, Kaylee Watson⁴ and Christine Yoshinaga-Itano^{1,2}

Abstract

Background: Oddball paradigms are frequently used to study auditory discrimination by comparing event-related potential (ERP) responses from a standard, high probability sound and to a deviant, low probability sound. Previous research has established that such paradigms, such as the mismatch response or mismatch negativity, are useful for examining auditory processes in young children and infants across various sleep and attention states. The extent to which oddball ERP responses may reflect subtle discrimination effects, such as speech discrimination, is largely unknown, especially in infants that have not yet acquired speech and language.

Results: Mismatch responses for three contrasts (non-speech, vowel, and consonant) were computed as a spectral-temporal probability function in 24 infants, and analyzed at the group level by a modified multidimensional scaling. Immediately following an onset gamma response (30–50 Hz), the emergence of a beta oscillation (12–30 Hz) was temporally coupled with a lower frequency theta oscillation (2–8 Hz). The spectral-temporal probability of this coupling effect relative to a subsequent theta modulation corresponds with discrimination difficulty for non-speech, vowel, and consonant contrast features.

Discussion: The theta modulation effect suggests that unexpected sounds are encoded as a probabilistic measure of surprise. These results support the notion that auditory discrimination is driven by the development of brain networks for predictive processing, and can be measured in infants during sleep. The results presented here have implications for the interpretation of discrimination as a probabilistic process, and may provide a basis for the development of single-subject and single-trial classification in a clinically useful context.

Conclusion: An infant's brain is processing information about the environment and performing computations, even during sleep. These computations reflect subtle differences in acoustic feature processing that are necessary for language-learning. Results from this study suggest that brain responses to deviant sounds in an oddball paradigm follow a cascade of oscillatory modulations. This cascade begins with a gamma response that later emerges as a beta synchronization, which is temporally coupled with a theta modulation, and followed by a second, subsequent theta modulation. The difference in frequency and timing of the theta modulations appears to reflect a measure of surprise. These insights into the neurophysiological mechanisms of auditory discrimination provide a basis for exploring the clinically utility of the MMR_{TF} and other auditory oddball responses.

Keywords: EEG, Auditory evoked potentials, Mismatch negativity, Speech discrimination, Infants, Sleep, Surprise

Background

Speech discrimination inherently requires a comparison of two or more different sounds, which, in EEG studies,

is typically assessed by some variant of an oddball response task [cf., 1, 2], and elicits an event-related potential (ERP) in the EEG signal. An oddball task includes the repeated presentation of two sounds¹ over many trials with some associated probability of which

*Correspondence: gilley@colorado.edu

¹ Institute of Cognitive Science, University of Colorado, Boulder, Boulder, CO, USA

Full list of author information is available at the end of the article

¹ Such tasks need not be limited to only two sounds; however, we limit the discussion here to a traditional oddball paradigm of two contrasting stimuli.

sound is selected during a trial. The sound with the highest probability of selection (i.e., the sound heard most often) is referred to as the “standard” or “frequent” stimulus, while the sound with the lowest probability is called the “deviant” or “rare” stimulus. If the task is meant to assess pre-perceptual processing, then the participant is usually asked to ignore the sounds and/or to attend to a separate task. Because these oddball responses can be assessed in the absence of a behavioral component, such a task might provide a solution to assessing discrimination in young infants.

Pre-perceptual oddball tasks are presumed to elicit an automatic, exogenous change-detection response [3]. For example, the mismatch negativity or mismatch response² (MMR/N) has been well defined in adults [4–6] and children [7–10], and has been correlated with behavioral discrimination [11, 12]. As the MMR/N is elicited passively and in the absence of directed attention [1, 13, 14], it has considerable potential as a measure of early information processing in infants [15]. Although considered a pre-attentive response, there has been debate on the modulatory effects of attention on the response [16]. Studies with adults attempt to control attention by having participants attend a visual stimulus during testing, but controlling attention in awake/alert infants may be more problematic. Conversely, studies in infants report stability in key ERP parameters (latency, amplitude, etc.) across various stages of alertness, including sleep [9, 17, 18]. Testing during sleep not only reduces the potential influence of attention effects, it has the added benefit of reducing data loss associated with movement artifacts.

During sleep, EEG in young children contains more low frequency power in the delta (2–4 Hz) and theta (4–8 Hz) frequency bands compared to older children and adults [19]. However, the exact contributions of these frequency bands also vary by sleep stage [20], and some infants (~35%) may exhibit transient theta-alpha (8–12 Hz) bursts in the EEG [21]. Despite such variance in sleep-stage contributions, auditory ERPs reveal typical auditory processing patterns during sleep and through the sleep-wake cycle [20]. However, detecting ERP differences in higher frequency bands, such as beta (12–30 Hz) and gamma (30–50 Hz) can be challenging given the smaller contributions to the EEG and unknown variability at those higher frequencies.

Several studies have now demonstrated theta amplitude modulations (AM) for frontal MMR/N components concurrent with theta phase modulations (PM) for temporal MMR/N components (i.e., as observed over temporal

cortex) when detecting a deviant stimulus [22–24]. These theta modulations also appear to follow a series of modulations in the gamma band (including high gamma, 60–300 Hz) and may be related to inhibitory changes in the alpha band [25]. Taken together, these previous results suggest a dynamical relationship between different neurophysiological mechanisms that contribute to the detection of a deviant sound. In the present study, we examine two hypotheses: (1) spectral-temporal features of auditory discrimination can be extracted from a “single-channel” montage that is ideal for clinical application in infants, and (2) these spectral-temporal features reflect the degree of difficulty for discriminating two sounds.

Methods

Participants

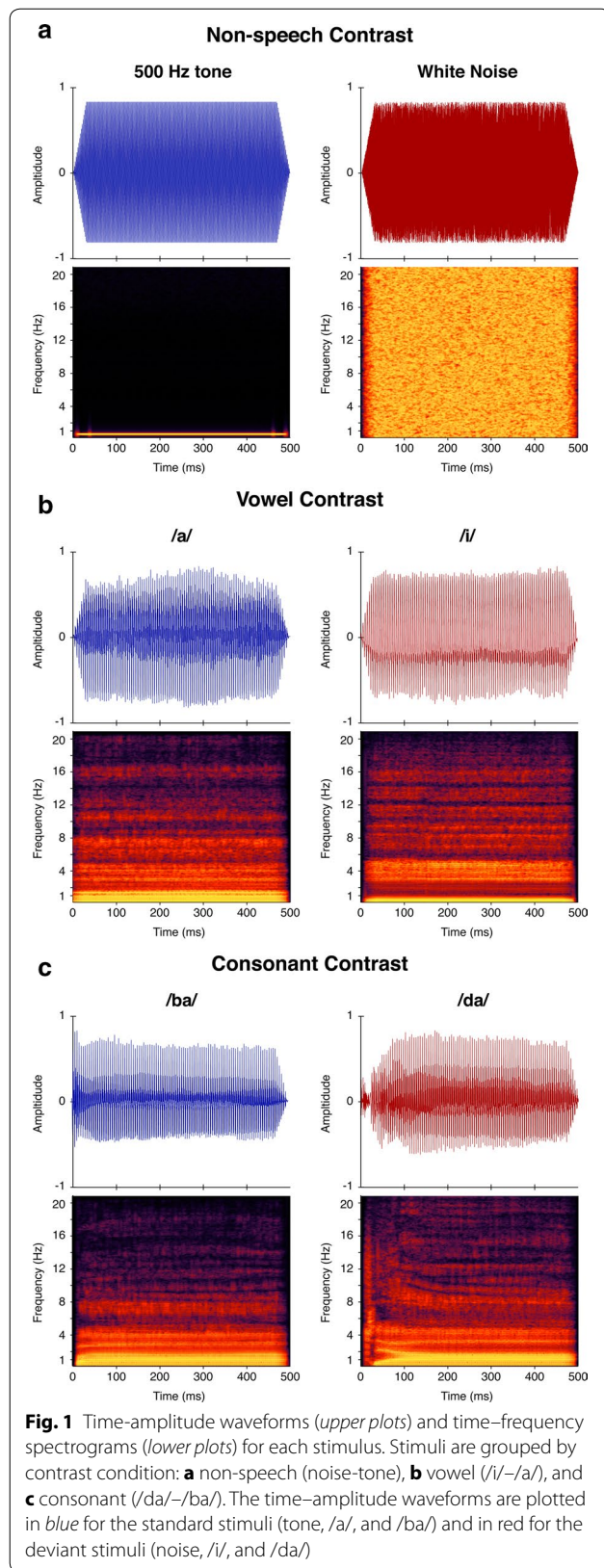
Participants for this study included 24 typically developing infants (10 female, 14 male) aged 1.0–3.9 months (mean = 2.6, SD = 0.82). All infants had passed their newborn hearing screening assessed by a click evoked auditory brainstem response (ABR) screening protocol. The informed consent form was fully executed prior to any study related activity, as approved by the Colorado Multiple Institutional Review Board.

Stimuli

Three stimulus pairs, or “contrasts” were presented in separate blocks using a standard auditory oddball paradigm with an inter-stimulus interval (ISI) of 1200 ms. This long interval increases the likelihood of identifying an auditory evoked response in young children [26, 27]. Stimuli were presented in pseudo-random order at a ratio of 85% *standard* to 15% *deviant*, with the constraint that deviant stimuli could not appear in succession. Approximately 600 trials (~510 standard, ~90 deviant) were collected for each block. All stimuli were normalized for RMS amplitude and presented from a single speaker in the sound field at a level of 70 dBA measured at the location of the infant’s head. All speech stimuli were edited to durations of 500 ms by replicating or cutting vowel cycles without disrupting the natural onset and offset of the sounds.

We chose three contrasts that represent increasing levels of difficulty and developmental emergence for auditory discrimination in young children; these include a non-speech contrast, a vowel contrast, and a consonant contrast (Fig. 1). The order of contrast presentation was randomly selected prior to each recording session. The non-speech contrast is considered the least difficult to discriminate in young children, and consisted of a 500 Hz pure tone (standard) and a white-noise burst (deviant) each 500 ms in duration with 30 ms linear ramping at both onset and offset. The vowel contrast consisted of

² While traditionally referred to as the mismatch negativity (MMN), infants exhibit a difference response with positive orientation, which is reflected by the more inclusive use of the mismatch response (MMR) to describe these responses.



two naturally produced vowel sounds /a/ (“ah”) and /i/ (“ee”) as the standard and deviant stimuli, respectively. This vowel contrast is considered more difficult to discriminate than the non-speech contrast, and is one of the earliest to emerge in behavioral discrimination tasks [28–30]. The consonant contrast consisted of two naturally spoken consonant–vowel (CV) sounds /ba/ (“bah”) and /da/ (“dah”) as the standard and deviant stimuli, respectively. This contrast is considered the most difficult of the three and tends to emerge later in behavioral discrimination tasks than the vowel contrast. The non-speech and vowel contrasts were completed for all 24 infants, but the consonant contrast was only completed for 17 of the 24 infants.

EEG procedure

Infants were placed in a comfortable bassinet or in a parent’s lap in a quiet, dim room to induce or aid sleeping during the test session. The rocker’s motion was not active during the EEG recordings, but was active during EEG preparation or during breaks if the infant appeared to be waking. Eleven Ag/AgCl electrodes were placed on the scalp according to the international 10–20 system (F5, Fz, F6, C5, Cz, C6, P5, Pz, P6, M1, and M2) and were referenced to the nasion. An additional bi-polar recording channel was placed on the lateral canthus of the right eye and referenced to the superior orbit to monitor eye movement and waking. Continuous EEG was recorded with a sampling rate of 1000 Hz and filtered from DC–100 Hz during each experimental block using a Synamps2 EEG amplifier (Compumedics-Neuroscan, Charlotte, NC).

EEG signal processing

All signal processing was conducted in Matlab R2015b (Mathworks, Natick, MA) using the Statistics toolbox with custom tools and the EEGLAB toolbox [31]. Prior to analysis of the experimental trials, the EEG data were filtered from 2 to 50 Hz (zero-phase, FIR, 24 dB/octave). This frequency range includes frequencies associated with early sensory and perceptual evoked potentials and with a high-pass filter above the range of the slow-wave activity during sleep. During the EEG recordings, sufficient EEG was captured before and after each block so that evoked responses were not distorted by the filter edges. Each channel (excluding the eye channel) was then re-referenced to the common average of all 11 data channels. Each contrast block was segmented into epochs from –500 to 1500 ms around each stimulus onset, and baseline corrected to the pre-stimulus interval. Trials with activity greater than 2.5 standard deviations from

the mean of the joint-probability distribution of trial amplitudes at each channel were considered as containing artifacts and were rejected from further analysis.

Channel selection

The first goal of this experiment was to determine an optimal recording montage for routine clinical application. We performed a spatial principal components analysis (PCA) on the bootstrapped difference estimate (n = 1001 bootstraps performed individually for each channel) between all standard and deviant trials. Difference estimates for each subject were normalized and concatenated for a group-level PCA. Components were sorted in descending order by the percentage of variance accounted for (pvaf) in the total data. The top components accounting for a cumulative variance of at least 90% were projected onto the channel space (Fig. 2a), and relative magnitudes were then computed as the sum of all trials. This results in a set of “loadings” for each EEG channel (Fig. 2b). Results of this analysis revealed three electrodes with consistently high loadings in all 24 participants: Cz, M1, and M2. Further, the loadings for each of the mastoid channels were negative while the more superior channels had positive loadings. This reflects the polarity inversion of the auditory dipole being above the axial plane of the mastoid electrodes and below the axial plane of the superior electrodes, and confirms the standard practice of recording auditory ERPs from Cz referenced to the mastoids. Taken together, we determined that Cz, referenced to linked mastoids (M1 + M2), would optimize differences in the experimental paradigm.

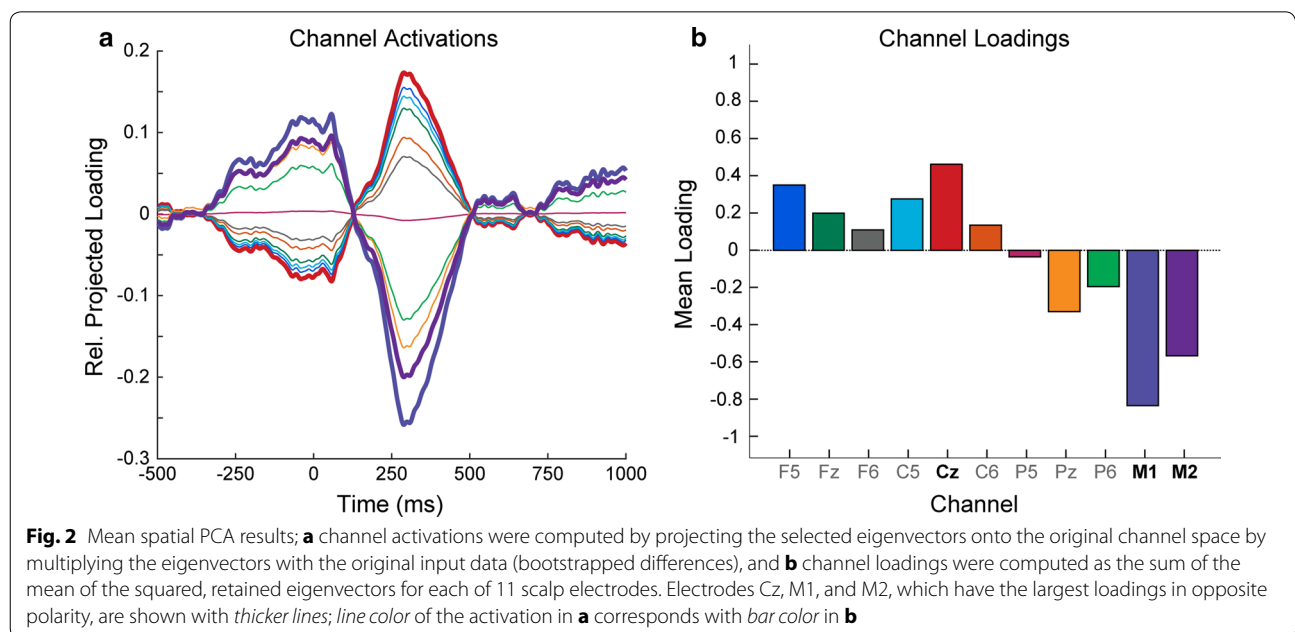
Therefore, the continuous data were re-referenced to the linked mastoids, and only the Cz channel was retained for further analysis.

Feature extraction

The second and primary goal of this experiment was to examine the spectral-temporal features of the mismatch response (MMR) in infants during sleep. We hypothesized that exogenous auditory processing during sleep would occur in separable EEG frequency bands, and that spectral-temporal features in these bands would reveal responses that indicate the detection of a deviant stimulus [15]. Although we ultimately seek to define such features for individuals, we focus this analysis on group-level effects to better understand the general processes underlying the MMR. To this end, we implemented a modified multi-dimensional scaling (MDS) of the time–frequency transformed EEG trials [32]. This analysis borrows from the DISTATIS method for the analysis of multiple distance matrices [33, 34]; where, here, distances refer to estimated differences in the deviant and standard responses in the time–frequency domain.

Pre-processing and time–frequency transformation

Each experimental block (i.e., contrast) from each subject was processed separately, and was defined as a 2-dimensional matrix of EEG voltages with size [I × J], where I is the number of trials and J is the number of time points in each trial. Each block was centered and whitened along the Ith dimension via PCA, retaining



all eigenvectors explaining at least 0.01% of the total variance. The continuous wavelet transform (CWT) was applied separately to each whitened trial using a 6-cycle Morlet wavelet with 128 log-spaced scales and with center frequencies from 1.94 to 48.40 Hz. The CWT results in a complex, 3-dimensional tensor of size $[I \times J \times K]$, where K is the number of scales ($K = 128$) in the transform, and each point includes both a real and imaginary component. After transformation, the time window for each trial was truncated to a range of -100 to 700 ms around the stimulus onset, and the pre-stimulus mean (-100 to 0 ms) was then subtracted from each trial. The time window truncation ensures that all time-frequency results are outside of the cone-of-influence; that is, that the data are not susceptible to edge artifacts. Trials with total squared magnitudes greater than 2.5 standard deviations from the mean were then rejected from the block.

Computing the MMR_{TF}

The probabilistic time-frequency mismatch response (MMR_{TF}) was defined as a probability of the complex magnitude of the estimated mean of the deviant (D) minus standard (S) trials being greater than zero:

$$MMR_{TF} \stackrel{\text{def}}{=} P((\overline{D - S}) > 0) \tag{1}$$

The estimated mean (D minus S bar) was computed by a bootstrap difference procedure ($n\text{-boots} = 1001$) that estimates both the mean difference and the probability of a difference being greater than 0 (i.e., deviant \neq standard). This method borrows from and extends the methods described for the integrated mismatch negativity (MMN_i) by Ponton et al. [35]. The estimated mean, M , is computed by randomly selecting one deviant trial and one standard trial (with replacement) during each bootstrap iteration and taking the difference, and is computed as:

$$M = \left\| \frac{1}{n} \sum_1^n d_{rd} - s_{rs} \right\|^2 \tag{2}$$

where n is the number of bootstraps (1001), d_{rd} is a randomly selected deviant trial, and s_{rs} is a randomly selected standard trial. The term, $\|\cdot\|^2$, indicates the squared complex modulus of the mean difference response. The estimated mean error, E , is computed by subtraction of any two randomly selected trials without regard to type (i.e., ignoring standard or deviant):

$$E = \left\| \frac{1}{n} \sum_1^n t_{ra} - t_{rb} \right\|^2 \tag{3}$$

where t_{ra} is any random trial from the entire block and t_{rb} is also any random trial from the entire block. The two results, M and E , are each real matrices of size $[K \times J]$, where J is the number of time points and K is the number of scales. To estimate the probability of a difference, M and E were vector normalized across scales, such that the squared sum of all points was equal to one for each scale:

$$MM_k = \frac{M_k}{\sqrt{\sum_{j=1}^J M_{j,k}^2 + \sum_{j=1}^J E_{j,k}^2}} \tag{4}$$

$$EE_k = \frac{E_k}{\sqrt{\sum_{j=1}^J M_{j,k}^2 + \sum_{j=1}^J E_{j,k}^2}} \tag{5}$$

where k is the K th scale, j is the the J th point, and MM and EE are the normalized means estimates. Therefore, the denominator term reads as the square root of the sum of all squared points in both MM and EE for scale k . This scale normalization has the effect that all difference points in the time-frequency plane are equated for the energy differences at each scale (e.g., low frequencies naturally have more power over time than higher frequencies). The joint cumulative density function (CDF) for all points in both MM and EE was then computed via kernel density estimation with automatic bandwidth selection [36] which we denote as:

$$C = CDF(MM \cup EE) \tag{6}$$

Finally, the probability function for $\overline{D - S}$ is defined by replacing values in MM with the probability of that value from C :

$$P((\overline{D - S}) > 0) \stackrel{\text{def}}{=} MM \rightarrow C \tag{7}$$

Therefore, Eq. 1, which defines the MMR_{TF} , is in turn defined by the joint probability of the true estimation (MM) and the error estimation (EE), and has a size of $[K \times J]$.

Group-level analysis

The aim of the group-level analysis was to extract and identify the spectral-temporal features that best explain differences between deviant and standard trials, and that may differentiate variances attributed to contrast difficulty. To achieve this aim, the MMR_{TF} results from each subject and contrast were treated as independent “studies” in a modified multi-dimensional scaling (MDS) analysis [32, 33].

Spectral and temporal cross-products matrices

Each MMR_{TF} was treated as a separate distance table and transformed into two cross-products matrices (as in MDS) for the spectral and temporal dimensions, respectively:

$$CP_F = MMR_{TF} * MMR_{TF}^T \tag{8}$$

$$CP_T = MMR_{TF}^T * MMR_{TF} \tag{9}$$

where CP_F is the cross-product matrix for the spectral (frequency) representation and CP_T represents the temporal (time) representation. The superscript, T, denotes the transpose.³ The centering matrix for each CP was defined as

$$\Xi = \mathbf{I} - \mathbf{I}\mathbf{m}^T \tag{10}$$

where \mathbf{I} is the identity matrix for CP with a size of $[n \times n]$, and where n is equal to the size of the dimension in CP (i.e., n is the number of scales represented in CP_F or the number of points represented in CP_T). The term $\mathbf{I}\mathbf{m}^T$ represents the “mass” contribution for each point in CP , where \mathbf{I} is a vector of ones with a size of $[n \times 1]$ and \mathbf{m} is mass vector of size $[n \times 1]$ and whose sum is equal to 1. We set each element of the mass vector to be equal, so each value of \mathbf{m} was equal to $1/n$. The superscript, T, denotes the transpose of \mathbf{m} . The centering matrix was then applied to each CP as:

$$\widetilde{CP}_F = -\frac{1}{2}\Xi_F CP_F \Xi_F^T \tag{11}$$

$$\widetilde{CP}_T = -\frac{1}{2}\Xi_T CP_T \Xi_T^T \tag{12}$$

where the subscripts F and T denote the spectral and temporal domains, respectively, and the superscript T denotes the transpose. Finally, each CP matrix is normalized by its first eigenvalue (λ^1) and denoted by the symbol \mathbf{S} :

$$\mathbf{S}_F = \lambda_F^{1-1} \times \widetilde{CP}_F \tag{13}$$

$$\mathbf{S}_T = \lambda_T^{1-1} \times \widetilde{CP}_{TT} \tag{14}$$

Joint DISTATIS

The goal of the modified DISTATIS analysis was to identify a set of weights, denoted \mathbf{W} , which define the best latent representation of the MMR_{TF} . In our case, we have two separated dimensions (spectral and temporal) represented as normalized cross-products matrices (\mathbf{S}_F and \mathbf{S}_T) for each MMR_{TF} . We therefore define two sets of weights, \mathbf{W}_F and \mathbf{W}_T , for each dimension. Each set of weights was computed via eigendecomposition of a compromise matrix, \mathbf{C} , for each dimension. The compromise matrix was computed as a weighted mean of all \mathbf{S} matrices of

the dimension. The mean was weighted in that separate means were first computed for each contrast condition (non-speech, vowel, and consonant), and the final compromise was computed as the mean of the individual condition compromises. Therefore, the condition compromise matrices were defined as:

$$C_F^d = \frac{1}{N} \sum_{n=1}^N \mathbf{S}_{F_n}^d \tag{15}$$

$$C_T^d = \frac{1}{N} \sum_{n=1}^N \mathbf{S}_{T_n}^d \tag{16}$$

where the superscript d refers to members of the contrast condition, and N is the number of members in that condition. In our case, N was equal to 24 for the non-speech and vowel contrasts, and 17 for the consonant contrasts. The final compromise for each dimension was computed as:

$$C_F = \frac{1}{D} \sum_{d=1}^D C_F^d \tag{17}$$

$$C_T = \frac{1}{D} \sum_{d=1}^D C_T^d \tag{18}$$

where D is the number of contrasts ($D = 3$) and d refers to the condition compromise matrices. That is, each final compromise is simply the mean of the condition compromises. Eigendecomposition of the final compromise matrices result in a set of eigenvectors, \mathbf{w} , and a corresponding set of eigenvalues, λ for each compromise. The eigenvalues can be used to determine the percentage of variance accounted for (pva_f) by each eigenvector (as in singular value decomposition), where

$$pva_f = \frac{\sqrt{\lambda}}{\sum \sqrt{\lambda}} \tag{19}$$

The final weights for each dimension were selected by applying a threshold of 0.01, such that only those eigenvectors explaining at least 1% of the total variance were retained. The retained eigenvectors were normalized such that the sum of the squared values was equal to one for each vector:

$$\mathbf{W}_F^U = \frac{w_F^U}{\sum_{u=1}^U \sqrt{w_F^{u2}}} \tag{20}$$

$$\mathbf{W}_T^V = \frac{w_T^V}{\sum_{v=1}^V \sqrt{w_T^{v2}}} \tag{21}$$

³ Note that for all equations a superscripted “T” refers to the transpose of a matrix or vector, whereas the subscripted “T” refers to the temporal dimension of the data.

where U and V refer to the number of retained eigenvectors in F and T , respectively; and \mathbf{W} denotes the final set of weights for the subscripted dimension. The first eigenvalue, λ^1 , for each dimension was used to compute the relative contribution (rc) of each dimension to explaining the total variance of the MMR_{TF} :

$$rc_F = \frac{\lambda_F^1}{\lambda_F^1 + \lambda_T^1} \tag{22}$$

$$rc_T = \frac{\lambda_T^1}{\lambda_F^1 + \lambda_T^1} \tag{23}$$

Data projection

Having identified a set of weights for each dimension we assessed the contribution of each MMR_{TF} to a *joint compromise projection* (\mathbb{P}) of the spectral and temporal dimensions, and each MMR_{TF} member to its respective condition compromise. In this way, we represent \mathbb{P} as a weighted sum of the MMR_{TF} dimension after a whitening transform by the weights, \mathbf{W} , as follows:

$$\mathbb{P}_F(MMR_{TF}) = \mathbf{W}_F \times \mathbf{W}_F^T \times MMR_{TF} \tag{24}$$

$$\mathbb{P}_T(MMR_{TF}) = \mathbf{W}_T \times \mathbf{W}_T^T \times MMR_{TF} \tag{25}$$

where $\mathbb{P}(MMR_{TF})$ refers to the projection of the MMR_{TF} into the dimension’s compromise space (via whitening). It is important to note that each individual MMR_{TF} was whitened by a set of weights computed from the group-level analysis. The joint projected compromise was then computed as:

$$\mathbb{P} = (\mathbb{P}_F(MMR_{TF}) * rc_F) + (\mathbb{P}_T(MMR_{TF}) * rc_T)^T \tag{26}$$

where $*$ denotes dot multiplication of each projection point with the dimension’s relative contribution, rc , from Eqs. 22 and 23. The superscript T refers to the transpose.

After projecting each MMR_{TF} , the condition-level compromise was computed as the mean of all members in the contrast condition:

$$\mathbf{g}^d = \frac{1}{M} \sum_{m=1}^M \mathbb{P}_m^d \tag{27}$$

where M is the number of MMR_{TF} responses in the contrast condition, d ($M = 24$ for non-speech and vowel contrasts, and $M = 17$ for the consonant contrast). Finally, the group-level joint compromise, \mathbb{G} , was computed as the mean of the condition-level compromise projections:

$$\mathbb{G} = \frac{1}{D} \sum_{d=1}^D \mathbf{g}_d \tag{28}$$

where D is the number of conditions ($D = 3$). The values in both \mathbb{G} and \mathbf{g} are normalized values between 0 and 1, which represent a *relative probability* (RP) of a difference response for each time–frequency point in the plane (i.e., relative to each other point in the plane).

Quasi-likelihood estimation

In order to quantify the likelihood that an RP value is significant *on average*, we derived a pseudo M-estimator from the joint compromise matrices. M-estimators define a broad class of statistical estimators that minimize functions of the data (e.g., least-squares estimation, and maximum likelihood estimation). In this case, each MMR_{TF} is defined as a probability function from Eq. 1, and the $\mathbb{P}(MMR_{TF})$ projections were computed by a PCA-based whitening function resulting in a normalized relative probability. Because PCA inherently minimizes these functions, we interpret the joint compromise as the “best representation” of this minimization. However, because we are combining more than one “best representation” (i.e., the weighted sum of two minimized functions over two separate dimensions), and each of these representations can contain more than two vectors (i.e., we allow each minimization function to exist in a multi-dimensional space), the representations do not satisfy all properties of a robust M-estimator. Further, combining the dimensions requires projecting the minimized data into an inflated, three-dimensional space with unit-less values, which can make interpretation difficult.

To resolve this issue of interpretation, we mapped the inflated projection, \mathbb{G} , onto a probability function of \mathbb{G} , derived from the three condition compromise projections. We first define a minimum bound in the time–frequency plane as the maximum value for each time–frequency point from each of the three compromise matrices:

$$\mathbf{B}_{ij} \stackrel{\text{def}}{=} \max_{ij \in D} \mathbf{g}_{ij}^d \tag{29}$$

where \mathbf{B} represents the minimum bound, and the max term indicates that each time–frequency point ij in \mathbf{B} is defined by the maximum value from each of the three condition compromise matrices, where the number of conditions is denoted by D and each condition in the set is denoted by d . The likelihood is then defined as

$$qLE \stackrel{\text{def}}{=} \mathbb{G} \rightarrow CDF(\mathbf{B})^2 \tag{30}$$

where qLE refers to a quasi-likelihood estimate, and takes on a value between 0 and 1, and $CDF(\mathbf{B})^2$ is the squared cumulative density function of the minimum bound, \mathbf{B} , as computed by kernel density estimation with automatic bandwidth selection [36]. We interpret the qLE as the likelihood that an observation in any joint projected

RP value represents a true difference between the deviant and standard responses relative to other observations in the time–frequency plane. The qLE is a single $[K \times J]$ matrix, which we essentially treat as a likelihood map of the relative probabilities for these experimental conditions; that is, it tells us where in the time–frequency plane we are most likely to observe a difference response for any of the three tested contrasts.

Given that the qLE represents a probability function, we sought to identify a threshold that can be used for feature extraction and group-level comparisons. Based on pilot analyses of these data, we found that a qLE threshold of 0.8 provided enough headroom to compare differences between extracted features, and was low enough to reveal consistently similar features between all participants. We suspect that this threshold may vary for different experimental procedures or populations. For our purposes, we applied a qLE threshold of 0.8 to each MMR_{TF} result. It is certainly possible that some features of interest may be overlooked with this threshold, however, results of this experiment provide some confirmation that this approach reveals consistent features. Future studies using a variety of experimental manipulations will be needed to fully resolve the optimal thresholding procedure.

Results

Group-level joint compromise, \mathbb{G}

The group-level joint compromise is shown in Fig. 3. To identify the spectral-temporal features, we set a threshold on the qLE at 0.8, which reveals three general features of interest in the time–frequency plane. The first prominent feature is a burst of activity that begins in the theta band and sweeps downward into the delta band over time. This sweeping burst has two centroids, or peaks, at the beginning and end of the downward sweep. The first centroid appears at 4.7 Hz with a peak latency of 66 ms after stimulus onset, and the second centroid appears at 2.2 Hz with a peak latency of 183 ms. The second feature appears as another downward sweep with slightly higher frequencies and later latencies than the first feature. As before, two centroids define the beginning and end of this sweep with the first centroid at 6.2 Hz with a peak latency of 215 ms and the second centroid at 2.7 Hz with a peak latency of 526 ms. These two features are denoted as theta-1 and theta-2. These theta responses correspond to the temporal phase modulated (PM) theta component (theta-1) and the frontal amplitude modulated (AM) theta component (theta-2) [24, 37, 38].

The third feature in the compromise appears as a large cloudy region spanning the beta (12–30 Hz) and gamma (30–50 Hz) bands. The temporal distribution of this gamma-beta cloud also appears to align with the temporal distribution of the theta-1 response, described above. This alignment is confirmed by comparing the temporal

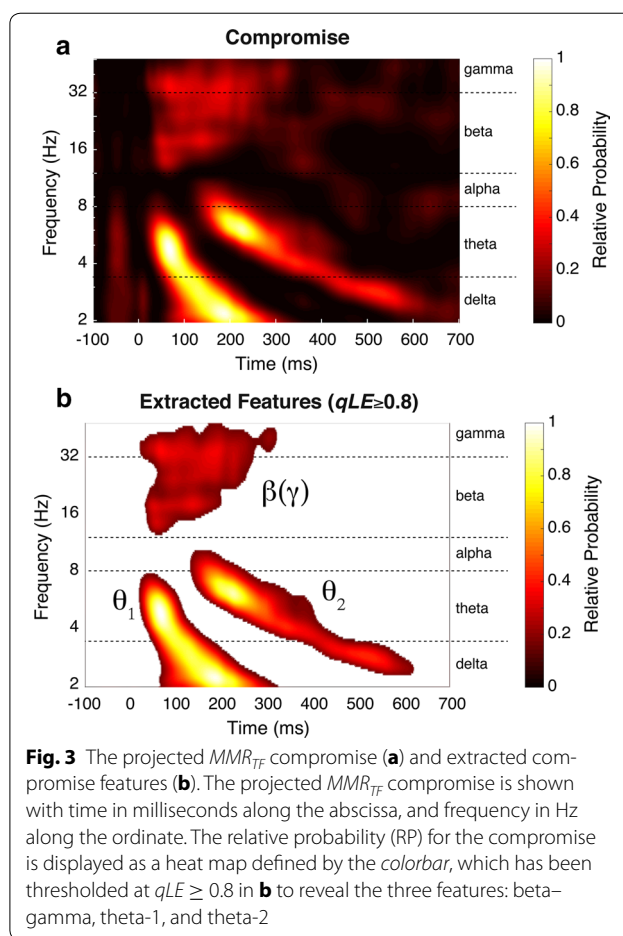


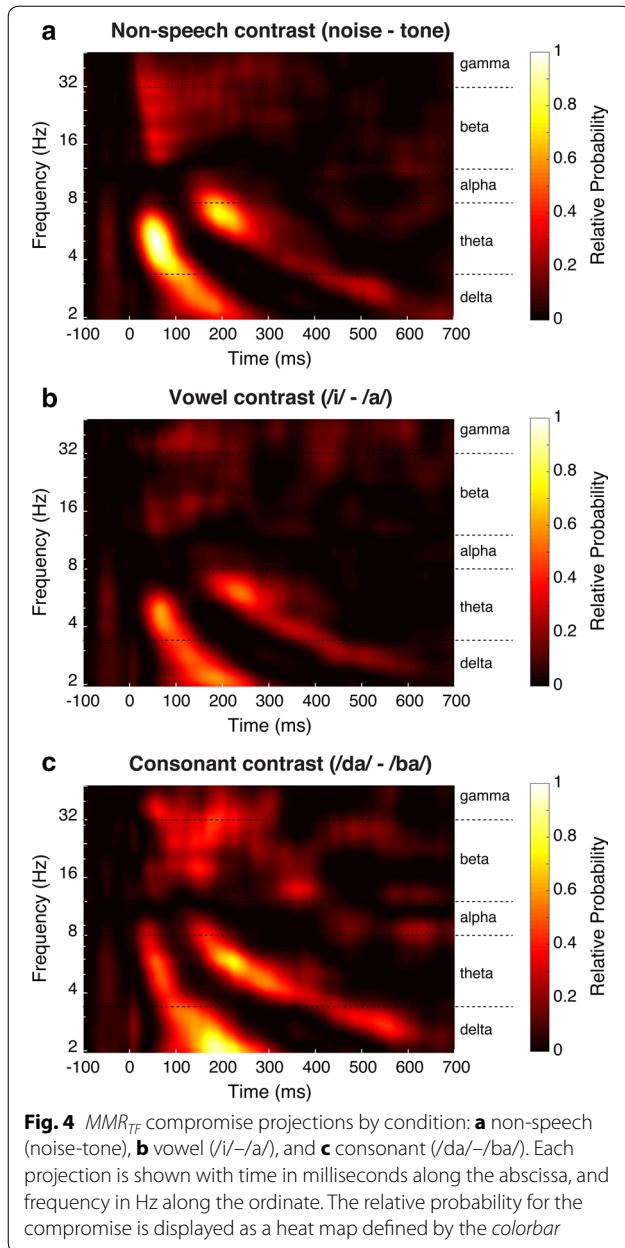
Fig. 3 The projected MMR_{TF} compromise (a) and extracted compromise features (b). The projected MMR_{TF} compromise is shown with time in milliseconds along the abscissa, and frequency in Hz along the ordinate. The relative probability (RP) for the compromise is displayed as a heat map defined by the colorbar, which has been thresholded at $qLE \geq 0.8$ in b to reveal the three features: beta–gamma, theta-1, and theta-2

distributions of the gamma-beta and theta-1 responses for each condition (see Fig. 5 and “Condition-level joint compromises, \mathbb{g}^d ” section, below). Such alignment corroborates evidence for a cross-frequency coupling (CFC) effect for information binding [39, 40].

Condition-level joint compromises, \mathbb{g}^d

The condition-level joint projections are shown in Fig. 4. The three features (theta-1, theta-2, and beta–gamma) of the group-level compromise are apparent in each of the condition-level projections, but with some notable differences upon visual inspection. For example, the theta-1 power appears to peak with a higher frequency and earlier latency for the non-speech and vowel contrasts compared to the consonant contrast.

The beta–gamma feature in the vowel contrast appears to be sparser and with less power than the other contrasts. There also appears to be an overall latency shift between the three features such that the non-speech contrast is earliest, followed by the vowel contrast, and then the consonant contrast. We sought to quantify these feature differences by comparing the centroids of each feature across contrast conditions.



We defined condition-feature centroids as the joint maximum of the temporal and spectral distributions for each feature. To derive these distributions, we created separate two-dimensional masks for each of the three features defined by $qLE \geq 0.8$. First, a mask H was created from the qLE values as:

$$H = \begin{cases} 0, & qLE < 0.8 \\ 1, & qLE \geq 0.8 \end{cases} \quad (31)$$

Separate masks were then created for each feature (d) by setting only those values bound within the feature to 1. Each feature mask (h^d) was then multiplied with

each condition-level projection (g^d), and the temporal (subscripted T) and spectral (subscripted F) distributions of the feature were then computed as the mean across the respective dimension:

$$y_T^d = \frac{1}{J} \sum_{j=1}^J (h_j^d * g_j^d) \quad (32)$$

$$y_F^d = \frac{1}{K} \sum_{k=1}^K (h_k^d * g_k^d) \quad (33)$$

where y denotes the mean RP distribution for the subscripted dimension and superscripted feature. Condition-feature distributions are shown in Figs. 5 and 6. Interestingly, the temporal distributions for the theta-1 and beta-gamma features are nearly identical in each of the conditions, which further supports a CFC effect between the theta-1 and gamma-beta responses.

Feature centroids for each condition were selected by conducting a search of the local maxima within each feature. This search revealed that the gamma-beta feature consists of two distinct maxima, a gamma and beta centroid, and that the theta-2 feature consisted of two distinct maxima corresponding to the approximate onset (theta-2) and offset (theta-2b) of the feature. The temporal and spectral distributions in Figs. 5 and 6 confirm that the gamma and beta centroids are distinct components, and that the theta-2b centroid appears to be a distinct component. The latency, frequency, and RP values for each centroid are listed in Table 1, and represented graphically in Fig. 7. We denote the second theta-2 centroid as theta-2b in Table 1, but it is not depicted in Fig. 7.

Discussion

Our aim was to examine and describe the spectral-temporal features of mismatch responses in the EEG of sleeping infants from a clinically viable electrode montage. The results of this study provide valuable insight into the neurophysiological mechanisms that underlie processes of prediction and discrimination in the developing brain. Below, we discuss the spectral-temporal features of these mechanisms, and describe a model for how these responses may reflect early speech discrimination.

The group-level compromise shown in Fig. 3 depicts, essentially, the experiment-wise probability of observing a spectral-temporal modulation during any given deviant trial. These spectral-temporal bounds are corroborated by previous findings of theta, alpha, beta, and gamma modulations at similar frequency ranges and latencies in auditory oddball responses. The present results suggest that

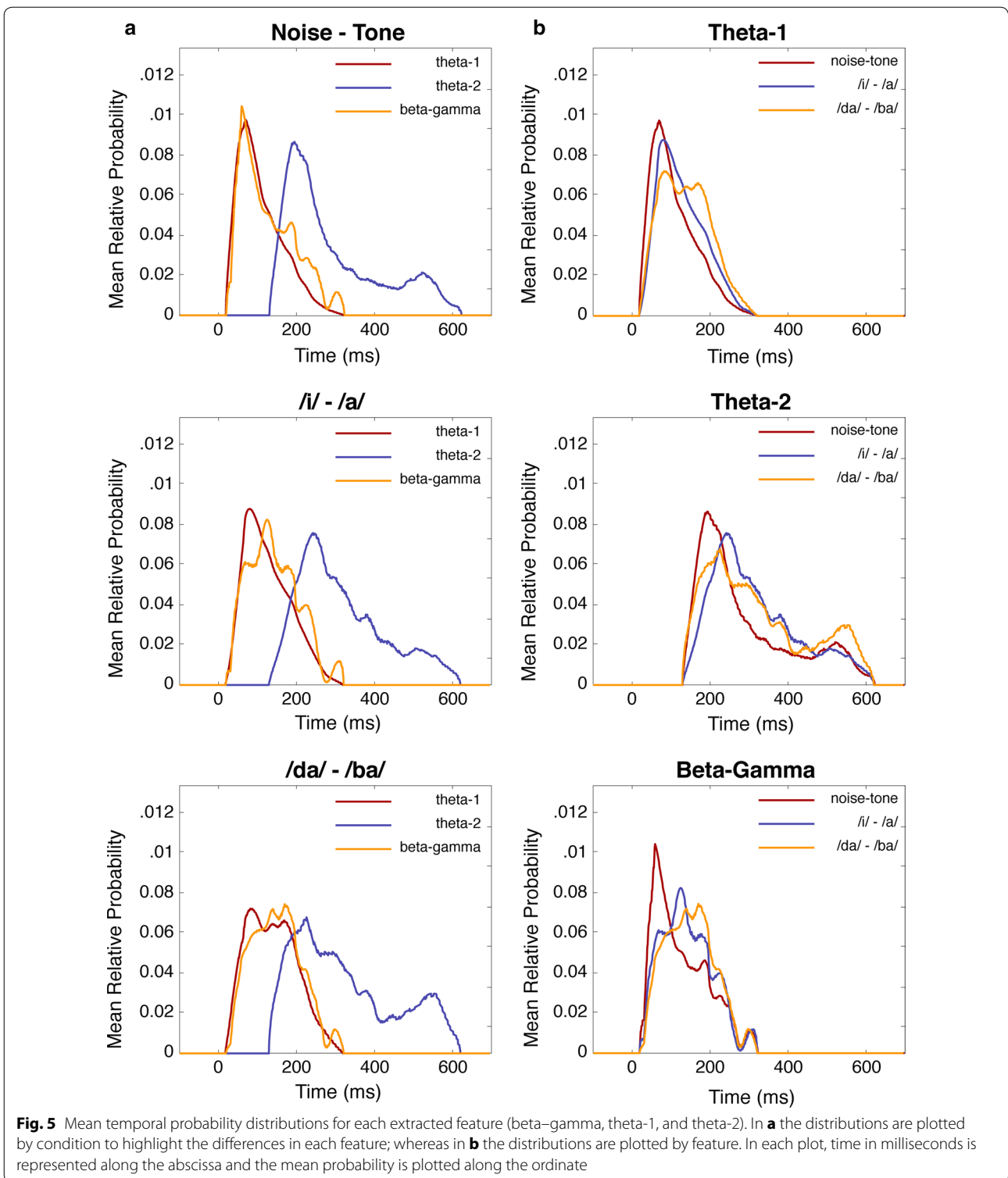


Fig. 5 Mean temporal probability distributions for each extracted feature (beta–gamma, theta-1, and theta-2). In **a** the distributions are plotted by condition to highlight the differences in each feature; whereas in **b** the distributions are plotted by feature. In each plot, time in milliseconds is represented along the abscissa and the mean probability is plotted along the ordinate

auditory oddball discrimination, as represented by the MMR_{TF} , occurs as a dynamical process between multiple neural generators spanning the delta, theta, beta, and

gamma frequency bands. We suspect the lack of an alpha response is due to sleep state, as alpha is generally associated with shifts in attentional processing while awake.

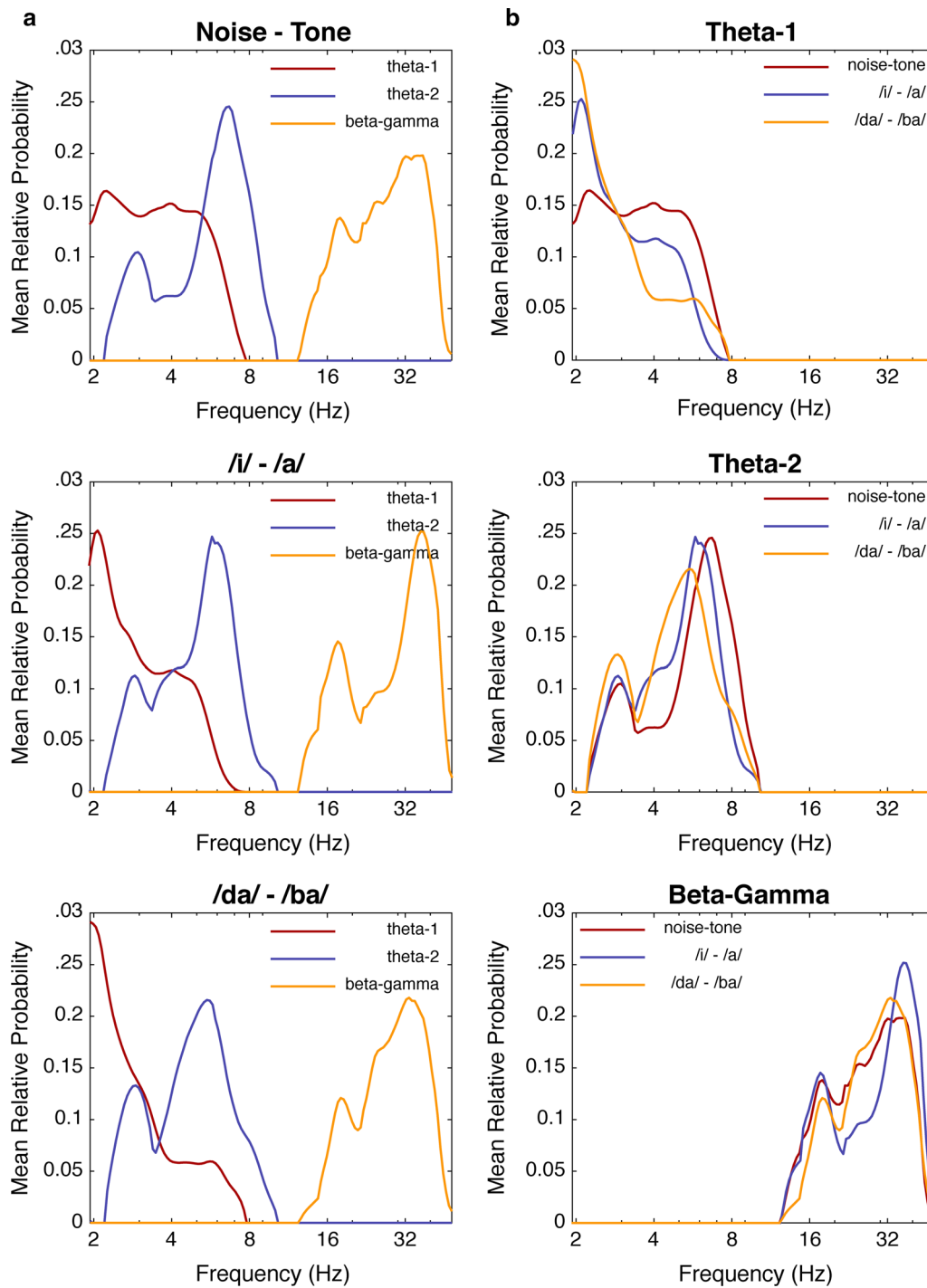


Fig. 6 Mean spectral probability distributions for each extracted feature (beta-gamma, theta-1, and theta-2). In **a** the distributions are plotted by condition to highlight the differences in each feature; whereas in **b** the distributions are plotted by feature. In each plot, frequency in Hz is represented along the abscissa and the mean probability is plotted along the ordinate

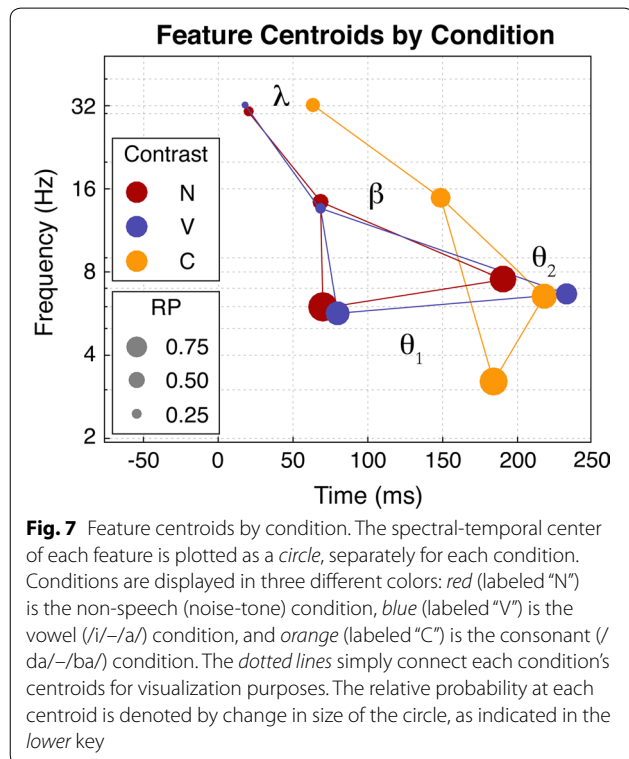
Discrimination and contrast difficulty

When comparing the condition-level compromise projections (Fig. 4), we did not anticipate the generally low

power in the responses to the vowel contrast (/i/-/a/) when compared to the non-speech (noise-tone) and consonant (/da/-/ba/) contrasts. Based on the ease of

Table 1 Centroid frequency, latency, and relative probability for each spectral-temporal feature (gamma, beta, theta-1, theta-2, and theta-2b) and each contrast condition: non-speech (noise-tone), vowel (/i/-/a/), and consonant (/da/-/ba/)

	Noise-tone	/i/-/a/	/da/-/ba/
<i>Gamma</i>			
Frequency (Hz)	33.94	35.71	35.71
Latency (ms)	22	20	59
Relative probability	0.25	0.09	0.38
<i>Beta</i>			
Frequency (Hz)	17.56	16.28	18.47
Latency (ms)	56	56	147
Relative probability	0.37	0.16	0.40
<i>Theta-1</i>			
Frequency (Hz)	5.07	4.70	2.09
Latency (ms)	58	69	187
Relative probability	1.00	0.62	0.92
<i>Theta-2</i>			
Frequency (Hz)	7.05	5.91	5.76
Latency (ms)	194	242	225
Relative probability	0.78	0.54	0.76
<i>Theta-2b</i>			
Frequency (Hz)	2.98	2.83	2.90
Latency (ms)	481	493	516
Relative probability	0.26	0.20	0.38

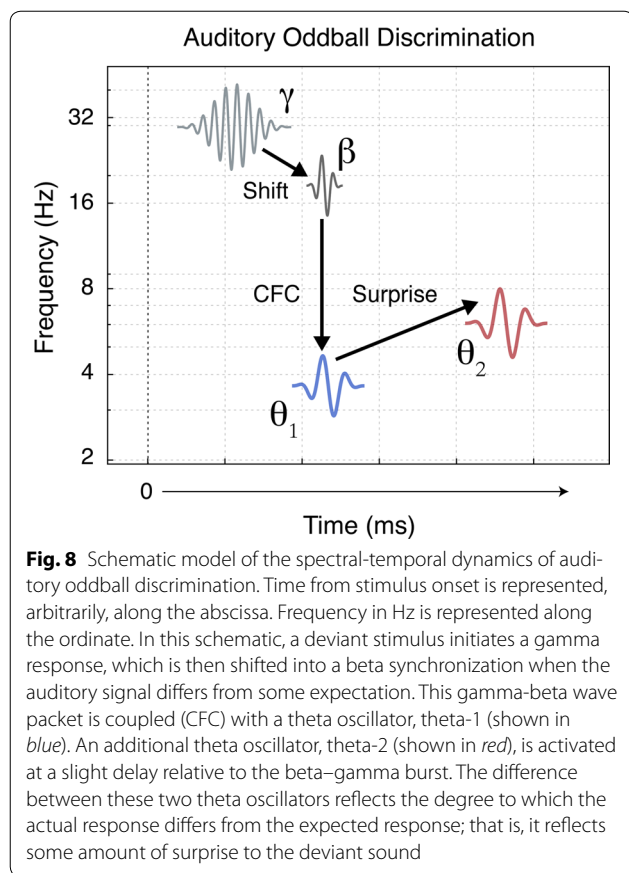


behavioral discrimination of /i/-/a/ [28–30] we predicted that the relative probabilities would be quite large and distinct. Rather, responses to the consonant and non-speech contrast were more robust than for the vowel contrast. There are several plausible explanations for this. First, the vowel contrast is composed of a spectral shift (i.e., a shift in vowel formant frequencies), whereas the consonant and non-speech contrasts are composed of both spectral shifts and rapid, aperiodic temporal shifts. It is possible that spectral differences are subtler than the more salient temporal differences, and thus elicit a smaller oddball response. However, this does not explain the ease of behavioral vowel discrimination.

A second explanation for the vowel difference led us to review the work by Saffran et al. [41–43] who have thoroughly examined various aspects of development of language-learning. Their research suggests that language learning occurs as a probabilistic function of sound and sound-pattern exposure. For example, infants learning English respond better to multi-syllabic patterns with stress on the first syllable, as those sounds are more likely to indicate a new word boundary. Functionally, a new word boundary indicates a greater likelihood of new information to be processed, which inherently requires more processing resources. Extended from this, processing resources for vowels, which span much longer time courses than consonants, require fewer resources to identify (relatively) slowly changing features such as formant shifts, which, probabilistically, do not alter the actual information content as extensively as a new word boundary. In this way, the smaller resource allocation for vowel processing would be observed as diffuse spectral power, which was observed here. The longer time course for processing the vowel information accounts for a greater probability of detecting a difference with relatively fewer resources expended, which may explain the ease of behavioral vowel discrimination.

Spectral-temporal dynamics

The centroids corresponding to each oscillatory component, as shown in Fig. 7, depict a spectral-temporal hierarchy for change detection [44–46], which is represented schematically in Fig. 8. This hierarchy suggests that auditory oddball discrimination is initiated as a gamma burst, which then likely emerges as a spatial amplitude modulation in the beta range (cf., [48]). This gamma–beta wave packet is temporally coupled with an oscillatory theta component, which is followed by a second theta component. The latency of the temporal coupling effect relative to the peak latency of a second theta oscillation reflects some degree of discrimination for the deviant sound.



Further, as the latency of the coupling effect increases the coupling frequency of the theta oscillation decreases to frequencies as low as 2 Hz (which was the lower limit of resolution for the spectral dimension). Taken together, this cascading process of modulation converges with evidence that oddball responses are driven by predictive processes that gate incoming sensory information by a measure of surprise (see footnote 3) [47–49].

Gamma and beta modulation

When applying a conservative threshold of $qLE \geq 0.8$, the responses in the gamma and beta bands appeared as a consolidated region (Fig. 3b), and are similar to the beta/gamma responses described by Isler et al. [15] and Stefanics et al. [50]. After separating the centroids for the beta and gamma responses (Fig. 7) we observed a distinct pattern of a gamma-beta shift that corroborates findings described in ERP experiments [51, 52] and from local field potentials and single-unit recordings of hippocampal and neocortical pyramidal cells in animal models [39, 53]. Results from animal models suggest that cells initially respond to a stimulus by releasing a burst of activity in the gamma range and then rapidly shift to a burst of slower activity in the beta range (with limitations; see

Traub et al. [53]). This gamma-beta shift may define a temporal window for long-range coherence between active components in a neural network, and may determine whether excitatory connections can synchronize at beta frequencies. For example, it is possible that local gamma oscillations act as a gating mechanism by limiting beta synchronization of gamma frequency responses to repeated, or *expected* stimulation. The presentation of a deviant sound would then, presumably, result in a larger gamma frequency response including responses from unexpected regions, and thus from uninhibited cells that are capable of beta synchronization. However, the gamma-beta shift does not, alone, describe the differences observed in an oddball response. For example, there must be something different about this particular shift than a shift that occurs during a standard response.

The speed at which the brain detects a difference in stimuli is quite remarkable, and is a process that begins as early as 20 ms after the stimulus onset. When detecting a deviant stimulus, change responses appear in the gamma band as early as 20 ms for easier contrasts and can be much later, ~60 ms, for more difficult contrasts, representing a wide range of variability for detection latency. The beta centroids appear slightly later at about 50–60 ms for easier contrasts, and can be much later at about 120–150 ms for more difficult contrasts. These beta centroids also correspond with the beta/gamma latencies reported by Isler et al. [15] and Stefanics et al. [50]. One explanation for this very rapid change detection is that these responses reflect neural mechanisms of pattern recognition and predictive processing; processes represented by gamma-beta shifts to changes in expectancy. Evidence for such predictive processing has been eloquently described in a review by Bendixen et al. [49]. In that review, the authors present a comprehensive model of predictive processing that (generally) includes a spectral-temporal hierarchy, such as that described by the results in this and other studies [44, 45, 52, 54]. To better understand this process, we must also account for the changes in lower frequency theta components that correspond to the detection of deviant sound.

Cross-frequency coupling

Another important feature of the MMR_{TF} is the concurrent timing of the gamma-beta and theta-1 modulations (see Fig. 5). Such concurrent timing between different frequency components is often described as a cross-frequency coupling (CFC) response (for review, see [39]). CFC is hypothesized as a mechanism for long-range synchrony between distinct, local generators of gamma activity. Schematically, we might think of the lower frequency theta waves as spatial modulators between network components operating in the gamma and beta

ranges (e.g., see [47]). With regard to deviance detection, the latency and coherence of CFC may act as a primary gating mechanism for incoming information, where some combination of latency and coupling strength provides a measure of surprise. The coupling latencies for each condition are represented in Fig. 7, which clearly shows a later CFC effect for the more difficult consonant contrast.

To make sense of this process, consider the simplicity of the basic oddball paradigm: one sound is repeated periodically, and a second sound is occasionally played instead; with this being repeated hundreds of times. If beta synchronization occurs via a gamma–beta shift as described above, then we might expect a larger gamma response to a deviant sound, followed by a noticeably larger beta synchronization. The emergence of beta synchronization occurs at some phase relative to a theta oscillator that itself is phase-locked to the expected rate of repetition [52]. It is plausible that the beta wave packet then perturbs this theta oscillator, which we observe as the CFC effect. As the original sound is repeated again, we should observe a habituation effect with each repetition, and we would expect that habituation to be most notable in beta frequency ranges and latencies. In the present study, we did not test for this habituation effect, but previous studies of habituation and refractoriness support this notion [26, 55].

Theta modulation

We might infer that the observation of a theta-1 modulation is driven by the CFC effect; that is, we observe a theta-1 perturbation concurrent with beta synchronization. However, this inference does not account for the observation of the theta-2 response, nor does the present study allow us to resolve this issue. One explanation is that the theta-2 response occurs as a release from local inhibition of beta synchronization by concurrent gamma oscillations, as described above. Another explanation is that a dedicated theta oscillator responds solely to perturbations of other theta oscillators that are actively processing incoming information. Yet another explanation is that the cause of the theta-2 response is simply not observed by these analysis methods. For example, because the difference estimates were computed in the complex time–frequency plane, we expect to observe spatial AM and PM effects as spectral perturbations, while FM and spectral power effects would be less noticeable. Other experimental methods, such as those that resolve synchronization/de-synchronization effects [51] may be better suited to resolving the catalyst of theta-2.

Fuentemilla, Marco-Pallarés et al. [22, 56] suggested that two distinct theta components, a frontal amplitude modulated component and a temporal phase modulated component, represent operations from distinct

neurophysiological mechanisms. Hsiao et al. [23] further suggested that these two theta components might characterize long-range phase synchrony within a temporo-frontal network for change detection. In the present results, theta-1 corresponds with the earlier temporal component and theta-2 with the later frontal component. Comparisons of the condition-level compromises show that the difference between easier and more difficult contrasts is represented by a spectral-temporal shift between these two theta components. We cannot say whether the time–frequency shift between the two theta components is a determinant of phase modulation, or “phase-resetting”, or a consequence thereof; however, we suggest that this time–frequency shift corresponds with such a mechanism.

Expectancy and surprise

Taken together, a dynamical process of expectancy and surprise⁴ can characterize deviance detection in an oddball paradigm. In this way, the entropy of an oscillating network characterizes the network’s expectation [47, 48]. When the network encounters some amount of surprise relative to the expected information, a perturbation of that network initiates the conditions for dissipating or transferring the unexpected information to higher-order processes. For example, the ability of local gamma responses to initiate beta synchronization may be dependent on some threshold for surprise. Further, this synchronization results in series of phase transitions to lower frequency components, as observed by the CFC effect subsequent to a gamma–beta shift. The actual amount of surprise might then be characterized by the spectral-temporal shift between theta-1 and theta-2. For example, the temporal distributions shown in Fig. 5a support this notion as the difference in the peaks of the theta distributions are closer together for more difficult contrasts, and include a greater likelihood of the response occurring in the lower delta range.

We suggest that the degree of difficulty for detecting a deviant stimulus is represented by a probability function that is constrained by the difference in both frequency and latency of the theta-1 and theta-2 components and by the latency and coupling strength of the beta–gamma and theta-1 components. Specifically, the process of deviance detection in an oddball paradigm can be characterized by a measure that is proportional to the surprise, S :

$$S \propto -\log\left(\frac{\partial F}{C\partial T}\right) \quad (34)$$

⁴ In information theory, surprise is defined by the negative logarithm of the expected information content (i.e., the entropy).

where ∂F is the probability of a difference in frequency between theta-1 and theta-2, and $C\partial T$ is the probability of a latency difference between theta-1 and theta-2 multiplied by a constant, C , which accounts for the latency of the CFC effect. This equation suggests that the spectral-temporal difference between theta-1 and theta-2 reflects the amount of information in the stimulus that was different than expected.

Experimental and clinical implications

The results reported here have implications for future experimental research utilizing an oddball paradigm, as well as for the clinical application of such methods. A key difference in our results and previous results is that spectral-temporal modulations are defined by a probability function instead of by measures of power or coherence. In this way, the MMR_{TF} represents a complex wave function that “collapses” upon observation [57]. Considering the MMR_{TF} as a wave function has implications for how we might interpret results from these experiments. If the wave function represents the probability of an observation, then the actual response during any given trial can appear anywhere within that function. Thus, for any single trial there is a significant amount of uncertainty as to whether such a response *did* occur, but after multiple observations we gather information to support some probability that the response *can* occur.

One interpretation of this probabilistic view is that any single-trial analysis of oddball responses must rely on some a priori information about the underlying system [58]. Fortunately, such an assumption provides a basis of support for implementing machine-learning methods for single-trial analyses. In this study, we chose the modified DISATIS algorithm because, mathematically, this approach acts as a precursor for kernel-based methods such as kernel-PCA, kernel-ICA, and support vector machines. Therefore, defining these responses as a probabilistic wave function provides considerable insight into how a machine learning approach might be implemented in future experiments. Unfortunately, such assumptions also infer a lower limit on the number of observations needed to achieve certainty for the probability function.

A key motivation for this research is the development of objective measures that might be used to assess auditory function in children with pathologies affecting the auditory system [15, 59]. For example, an infant identified with hearing loss must be fit with amplification (e.g., hearing aids) that is tuned to optimize information processing in certain frequency bands; for example, the speech frequencies. Whether an infant can discriminate two speech sounds depends on how well

the amplification is tuned, which has direct implications for language-learning [41, 60]. An objective measure such as the MMR_{TF} might provide a means to assess whether an infant has adequate access to the speech frequencies. From a clinical perspective, we wish to determine whether an infant *can* discriminate between two speech sounds. As another example, children with auditory processing disorders exhibit deficits in beta coupling observed as shifts in beta synchronization frequencies [52, 54]. A measure such as the MMR_{TF} might provide a means to better classify these deficits, or to provide real-time feedback during assessment. In this case, we wish to determine whether a child *did* discriminate between two sounds. Of course, to better understand these potential applications, further research including these populations will be necessary.

Clinical use of an oddball paradigm such as the MMR/N is often met with skepticism over the validity of results at the individual level [15, 61]. One source of such skepticism is the large variability in the latencies of MMR/N responses analyzed via traditional signal averaging. Indeed the spectral-temporal proximity of the theta-1 and theta-2 responses likely account for much of this variability, as these two components would appear as a mixture in the averaged ERP (cf., [62]). Further, the variability of the response onset—the CFC effect—contributes to the temporal overlap and subsequent smearing in the averaged response. Results reported here support the notion that time–frequency analyses improve the likelihood of detecting and classifying these responses within individuals.

Conclusions

Even during sleep, an infant’s brain is processing information about the environment and performing computations about that information in an unconscious state. Moreover, these computations reflect subtle differences in acoustic feature processing that are necessary for language-learning. Results from this study suggest that brain responses to deviant sounds in an oddball paradigm follow a cascade of oscillatory modulations. This cascade begins with a gamma response that later emerges as a beta synchronization, which is temporally coupled with a theta modulation, and followed by a second, subsequent theta modulation. The difference in frequency and timing of the theta modulations appears to reflect a measure of surprise; that is, it provides a measure of the error between the information that was expected, and the information that was actually received. These insights into the neurophysiological mechanisms of auditory discrimination provide a basis for exploring the clinical utility of the MMR_{TF} and other auditory oddball responses.

Abbreviations

AM: amplitude modulation; CFC: cross-frequency coupling; CWT: continuous wavelet transform; EEG: electroencephalography; ERP: event related potential; FM: frequency modulation; MDS: multidimensional scaling; MMR_{TF} : probabilistic time–frequency mismatch response; MMR/N: mismatch response/negativity; PCA: principal components analysis; PM: phase modulation; pvaf: percentage of variance accounted for; qLE: quasi-likelihood estimate.

Authors' contributions

KU and PMG contributed equally to this manuscript. KU conceived the experiments and led the experimental design. CYI assisted with stimulus development and the theoretical model of discrimination with minor input from PMG. KU, PMG, and KW conducted the EEG experiments. PMG contributed the analysis tools and conducted the EEG analysis. KU and PMG wrote the manuscript. All authors contributed edits to the manuscript, and read and approved the final manuscript.

Author details

¹ Institute of Cognitive Science, University of Colorado, Boulder, Boulder, CO, USA. ² Marion Downs Center, Denver, CO, USA. ³ University of Colorado School of Medicine, Aurora, CO, USA. ⁴ Department of Speech, Language, and Hearing Sciences, University of Colorado, Boulder, Boulder, CO, USA.

Acknowledgements

We gratefully acknowledge the feedback and input from Drs. Randy Ross and Sharon Hunter during the early stages of this experiment. During the revision stages of this manuscript, Dr. Ross passed away unexpectedly. Dr. Ross was a great mentor, collaborator, and friend, and his insights and contributions to our field of work will be greatly missed.

Competing interests

The authors declare that they have no competing interests.

Availability of data and materials

The data used for these analyses cannot be made publically available due to restrictions in our human subjects protocol, and because these data represent a subset of data from a larger, longitudinal study. Data may be made publically available at the completion of that study, dependent on further protocol restrictions. The software used to execute the MMR_{TF} computations and to perform the group-level analysis is available online in the MathWorks File-Exchange repository: *Project name:* AORtools v1.0. *Project home page:* <http://www.mathworks.com/matlabcentral/fileexchange/56138-aortools>. *Archived version:* 1.0. *Operating system:* Platform independent (developed with Mac OSX 10.10.5). *Requirements:* Matlab version R2014b or higher. *License:* BSD License.

Consent for publication

All data and results reported in this manuscript are presented at the group-level; i.e., no individual or personally identifying information is contained in this report, and individual consent for data to be included for analysis was obtained as part of the informed consent procedure.

Ethics approval and consent to participate

This research was approved by the Colorado Multiple Institutional Review Board: COMIRB 13-0342. The informed consent form was fully executed prior to any study related activity.

Funding

This research was supported by the National Institute on Deafness and Other Communication Disorders to author KU (NIDCD K23DC013583); and contents of this manuscript were developed under a grant from the National Institute on Disability, Independent Living, and Rehabilitation Research (NIDILRR Grant Number 90RE5020) to authors PMG and CYI. NIDILRR is a Center within the Administration for Community Living (ACL), Department of Health and Human Services (HHS). The contents of this research manuscript do not necessarily represent the policy of NIDILRR, ACL, HHS, and you should not assume endorsement by the Federal Government.

Received: 25 March 2016 Accepted: 9 March 2017

Published online: 22 March 2017

References

- Näätänen R, Jacobsen T, Winkler I. Memory-based or afferent processes in mismatch negativity (MMN): a review of the evidence. *Psychophysiology*. 2005;42:25–32.
- Näätänen R, Paavilainen P, Rinne T, Alho K. The mismatch negativity (MMN) in basic research of central auditory processing: a review. *Clin Neurophysiol*. 2007;118:2544–90.
- Näätänen R. Attention and brain function. Hillsdale: Lawrence Erlbaum; 1992.
- Näätänen R, Alho K. Mismatch negativity—a unique measure of sensory processing in audition. *Int J Neurosci*. 1995;80:317–37.
- Näätänen R, Alho K. Mismatch negativity—the measure for central sound representation accuracy. *Audiol Neurootol*. 1997;2:341–53.
- Näätänen R, Winkler I. The concept of auditory stimulus representation in cognitive neuroscience. *Psychol Bull*. 1999;125:826–59.
- Cheour M, Alho K, Ceponiene R, Reinikainen K, Sainio K, Pohjavuori M, Aaltonen O, Naatanen R. Maturation of mismatch negativity in infants. *Int J Psychophysiol*. 1998;29:217–26.
- Kushnerenko E, Ceponiene R, Balan P, Fellman V, Naatanen R. Maturation of the auditory change detection response in infants: a longitudinal ERP study. *NeuroReport*. 2002;13:1843–8.
- Kushnerenko E, Winkler I, Horvath J, Naatanen R, Pavlov I, Fellman V, Huotilainen M. Processing acoustic change and novelty in newborn infants. *Eur J Neurosci*. 2007;26:265–74.
- Mueller JL, Friederici AD, Mannel C. Auditory perception at the root of language learning. *Proc Natl Acad Sci USA*. 2012;109:15953–8.
- Kraus N, Micco AG, Koch DB, McGee T, Carrell T, Sharma A, Wiet RJ, Weingarten CZ. The mismatch negativity cortical evoked potential elicited by speech in cochlear-implant users. *Hear Res*. 1993;65:118–24.
- Javitt DC, Shelley AM, Ritter W. Associated deficits in mismatch negativity generation and tone matching in schizophrenia. *Clin Neurophysiol*. 2000;111:1733–7.
- Paavilainen P, Tiitinen H, Alho K, Näätänen R. Mismatch negativity to slight pitch changes outside strong attentional focus. *Biol Psychol*. 1993;37:23–41.
- Leppänen PHT, Lyytinen H. Auditory event-related potentials in the study of developmental language-related disorders. *Audiol Neurootol*. 1997;2:308–40.
- Isler JR, Tarullo AR, Grieve PG, Housman E, Kaku M, Stark RI, Fifer WP. Toward an electrocortical biomarker of cognition for newborn infants. *Dev Sci*. 2012;15:260–71.
- Sussman ES. A new view on the MMN and attention debate. *J Psychophysiol*. 2007;21:164–75.
- Kushnerenko E, Ceponiene R, Balan P, Fellman V, Huotilainen M, Näätänen R. Maturation of the auditory event-related potentials during the first year of life. *NeuroReport*. 2002;13:47–51.
- Kushnerenko E, Ceponiene R, Fellman V, Huotilainen M, Winkler I. Event-related potential correlates of sound duration: similar pattern from birth to adulthood. *NeuroReport*. 2001;12:3777–81.
- Kurth S, Achermann P, Rusterholz T, Lebourgeois MK. Development of brain EEG connectivity across early childhood: does sleep play a role? *Brain Sci*. 2013;3:1445–60.
- Portas CM, Krakow K, Allen P, Josephs O, Armony JL, Frith CD. Auditory processing across the sleep-wake cycle. *Neuron*. 2000;28:991–9.
- Hayakawa F, Watanabe K, Hakamada S, Kuno K, Aso K. Fz theta/alpha bursts: a transient EEG pattern in healthy newborns. *Electroencephalogr Clin Neurophysiol*. 1987;67:27–31.
- Fuentemilla L, Marco-Pallarés J, Münte TF, Grau C. Theta EEG oscillatory activity and auditory change detection. *Brain Res*. 2008;1220:93–101.
- Hsiao FJ, Cheng CH, Liao KK, Lin YY. Cortico-cortical phase synchrony in auditory mismatch processing. *Biol Psychol*. 2010;84:336–45.
- Ko D, Kwon S, Lee G, Im H, Kim H, Jung K. Theta oscillation related to the auditory discrimination process in mismatch negativity: oddball versus control paradigm. *J Clin Neurool*. 2012;8:35–42.
- Edwards E, Soltani M, Kim W, Dalal SS, Nagarajan SS, Berger MS, Knight RT. Comparison of time-frequency responses and the event-related potential to auditory speech stimuli in human cortex. *J Neurophysiol*. 2009;102:377–86.

26. Gilley PM, Sharma A, Dorman M, Martin K. Developmental changes in refractoriness of the cortical auditory evoked potential. *Clin Neurophysiol*. 2005;116:648–57.
27. Gilley PM, Sharma A, Dorman M, Martin K. Abnormalities in central auditory maturation in children with language-based learning problems. *Clin Neurophysiol*. 2006;117:1949–56.
28. Uhler KM, Baca R, Dudas E, Fredrickson T. Refining stimulus parameters in assessing infant speech perception using visual reinforcement infant speech discrimination: sensation level. *J Am Acad Audiol*. 2015;26:807–14.
29. Eisenberg LS, Martinez AS, Boothroyd A. Auditory-visual and auditory-only perception of phonetic contrasts in children. *Volta Rev*. 2003;103:327–46.
30. Boothroyd A. Auditory perception of speech contrasts by subjects with sensorineural hearing loss. *J Speech Hear Res*. 1984;27:134–44.
31. Delorme A, Makeig S. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J Neurosci Methods*. 2004;134:9–21.
32. Abdi H. Metric multidimensional scaling (MDS): analyzing distance matrices multidimensional scaling: eigen-analysis of a distance matrix. In: Salkind NJ, editor. *Encyclopedia of measurement and statistics*. Thousand Oaks: Sage; 2007. p. 598–605.
33. Abdi H, Valentin D. How to analyze multiple distance matrices. In: Salkind NJ, editor. *Encyclopedia of measurement and statistics*. Thousand Oaks: Sage; 2007. p. 284–90.
34. Abdi H, Valentin D, Chollet S, Chrea C. Analyzing assessors and products in sorting tasks: DISTATIS, theory and applications. *Food Qual Preference*. 2007;18:627–40.
35. Ponton CW, Don M, Eggermont JJ, Kwong B. Integrated mismatch negativity (MMN(i)): a noise-free representation of evoked responses allowing single-point distribution-free statistical tests. *Electroencephalogr Clin Neurophysiol Evoked Potentials*. 1997;104:143–50.
36. Botev ZI, Grotowski JF, Kroese DP. Kernel density estimation via diffusion. *Ann Stat*. 2010;38:2916–57.
37. Rinne T, Alho K, Ilmoniemi RJ, Virtanen J, Näätänen R. Separate time behaviors of the temporal and frontal mismatch negativity sources. *Neuroimage*. 2000;12:14–9.
38. Choi JW, Lee JK, Ko D, Lee GT, Jung KY, Kim KH. Fronto-temporal interactions in the theta-band during auditory deviant processing. *Neurosci Lett*. 2013;548:120–5.
39. Buzsáki G, Wang X-J. Mechanisms of gamma oscillations. *Annu Rev Neurosci*. 2012;35:203–25.
40. Belluscio MA, Mizuseki K, Schmidt R, Kempter R, Buzsáki G. Cross-frequency phase-phase coupling between θ and γ oscillations in the hippocampus. *J Neurosci*. 2012;32:423–35.
41. Saffran J, Hauser M, Seibel R, Kapfhamer J, Tsao F, Cushman F. Grammatical pattern learning by human infants and cotton-top tamarin monkeys. *Cognition*. 2008;107:479–500.
42. Saffran JR, Pollak SD, Seibel RL, Shkolnik A. Dog is a dog is a dog: Infant rule learning is not specific to language. *Cognition*. 2007;105:669–80.
43. Saffran JR, Reeck K, Niebuhr A, Wilson D. Changing the tune: the structure of the input affects infants' use of absolute and relative pitch. *Dev Sci*. 2005;8:1–7.
44. Lakatos P, Shah AS, Knuth KH, Ulbert I, Karmos G, Schroeder CE. An oscillatory hierarchy controlling neuronal excitability and stimulus processing in the auditory cortex. *J Neurophysiol*. 2005;94:1904–11.
45. Kiebel SJ, Daunizeau J, Friston KJ. A hierarchy of time-scales and the brain. *PLoS Comput Biol*. 2008;4:e1000209.
46. Ahissar M, Nahum M, Nelken I, Hochstein S. Reverse hierarchies and sensory learning. *Philos Trans R Soc Lond B Biol Sci*. 2009;364:285–99.
47. Freeman WJ, Vitiello G. Nonlinear brain dynamics as macroscopic manifestation of underlying many-body field dynamics. *Phys Life Rev*. 2006;3:93–118.
48. Friston K. The free-energy principle: a rough guide to the brain? *Trends Cogn Sci*. 2009;13:293–301.
49. Bendixen A, SanMiguel I, Schröger E. Early electrophysiological indicators for predictive processing in audition: a review. *Int J Psychophysiol*. 2012;83:120–31.
50. Stefanics G, Håden G, Huotilainen M, Balazs L, Sziller I, Beke A, Fellman V, Winkler I. Auditory temporal grouping in newborn infants. *Psychophysiology*. 2007;44:697–702.
51. Pantev C. Evoked and induced gamma-band activity of the human cortex. *Brain Topogr*. 1995;7:321–30.
52. Gilley PM, Sharma M, Purdy SC. Oscillatory decoupling differentiates auditory encoding deficits in children with listening problems. *Clin Neurophysiol*. 2016;127:1618–28.
53. Traub RD, Whittington MA, Buhl EH, Jefferys JG, Faulkner HJ. On the mechanism of the $\gamma \rightarrow \beta$ frequency shift in neuronal oscillations induced in rat hippocampal slices by tetanic stimulation. *J Neurosci*. 1999;19:1088–105.
54. Gilley P, Walker N, Sharma A. Abnormal oscillatory neural coupling in children with language-learning problems and auditory processing disorder. *Semin Hear*. 2014;1:15–26.
55. Budd T, Barry RJ, Gordon E, Rennie C, Michie P. Decrement of the N1 auditory event-related potential with stimulus repetition: habituation vs. refractoriness. *Int J Psychophysiol*. 1998;31:51–68.
56. Fuentemilla L, Marco-Pallarés J, Grau C. Modulation of spectral power and of phase resetting of EEG contributes differentially to the generation of auditory event-related potentials. *Neuroimage*. 2006;30:909–16.
57. Heisenberg W. Über den anschaulichen Inhalt der quantentheoretischen Kinematik und Mechanik [The actual content of quantum theoretical kinematics and mechanics]. *Zeitschrift für Phys [Translated NASA Tech Doc TM-77379, 1983] 1927;43:172–98*.
58. Lieder F, Daunizeau J, Garrido MI, Friston KJ, Stephan KE. Modelling trial-by-trial changes in the mismatch negativity. *PLoS Comput Biol*. 2013;9:e1002911.
59. Bishop DVM. Using mismatch negativity to study central auditory processing in developmental language and literacy impairments: where are we, and where should we be going? *Psychol Bull*. 2007;133:651–72.
60. Estes KG, Evans JL, Alibali MW, Saffran JR. Can infants map meaning to newly segmented words? *Psychol Sci*. 2007;18:254–60.
61. Bishop DVM, Hardiman MJ. Measurement of mismatch negativity in individuals: a study using single-trial analysis. *Psychophysiology*. 2010;47:697–705.
62. Bishop DVM, Hardiman M, Barry J. Lower-frequency event-related desynchronization: a signature of late mismatch responses to sounds, which is reduced or absent in children with specific language impairment. *J Neurosci*. 2010;30:15578–84.

Submit your next manuscript to BioMed Central
and we will help you at every step:

- We accept pre-submission inquiries
- Our selector tool helps you to find the most relevant journal
- We provide round the clock customer support
- Convenient online submission
- Thorough peer review
- Inclusion in PubMed and all major indexing services
- Maximum visibility for your research

Submit your manuscript at
www.biomedcentral.com/submit

