Do Auditory Mismatch Responses Differ Between Acoustic Features?

HyunJung An[1], Shing Ho Kei[1], Ryszard Auksztulewicz[1,2*†] and Jan W. H. Schnupp[1*†]

[1] Department of Neuroscience, City University of Hong Kong, Kowloon, Hong Kong, [2] Department of Neuroscience, Max Planck Institute for Empirical Aesthetics, Frankfurt, Germany

Mismatch negativity (MMN) is the electroencephalographic (EEG) waveform obtained by subtracting event-related potential (ERP) responses evoked by unexpected deviant stimuli from responses evoked by expected standard stimuli. While the MMN is thought to reflect an unexpected change in an ongoing, predictable stimulus, it is unknown whether MMN responses evoked by changes in different stimulus features have different magnitudes, latencies, and topographies. The present study aimed to investigate whether MMN responses differ depending on whether sudden stimulus change occur in pitch, duration, location or vowel identity, respectively. To calculate ERPs to standard and deviant stimuli, EEG signals were recorded in normal-hearing participants ($N = 20$; 13 males, 7 females) who listened to roving oddball sequences of artificial syllables. In the roving paradigm, any given stimulus is repeated several times to form a standard, and then suddenly replaced with a deviant stimulus which differs from the standard. Here, deviants differed from preceding standards along one of four features (pitch, duration, vowel or interaural level difference). The feature levels were individually chosen to match behavioral discrimination performance. We identified neural activity evoked by unexpected violations along all four acoustic dimensions. Evoked responses to deviant stimuli increased in amplitude relative to the responses to standard stimuli. A univariate (channel-by-channel) analysis yielded no significant differences between MMN responses following violations of different features. However, in a multivariate analysis (pooling information from multiple EEG channels), acoustic features could be decoded from the topography of mismatch responses, although at later latencies than those typical for MMN. These results support the notion that deviant feature detection may be subserved by a different process than general mismatch detection.

Keywords: electroencephalography, mismatch negativity, predictive coding, auditory processing, multivariate decoding

INTRODUCTION

Neural activity is typically suppressed in response to expected stimuli and enhanced following novel stimuli (Carbajal and Malmierca, 2018). This effect is often summarized as a mismatch response, calculated by subtracting the neural response waveform to unexpected deviant stimuli from the response to expected standard stimuli. Auditory deviance detection has been associated with a human auditory-evoked potential, the mismatch negativity, occurring at about 150–250 ms from sound change onset (Naatanen, 2007; Garrido et al., 2008). The principal neural sources of the MMN are thought to be superior temporal regions adjacent to the primary auditory cortex, as

well as frontoparietal areas (Doeller et al., 2003; Chennu et al., 2013). Initially, the MMN was interpreted as a correlate of pre-attentive encoding of physical features between standard and deviant sounds (Doeller et al., 2003). However, more recent studies have led to substantial revisions of this hypothesis, and currently, the most widely accepted explanation of the MMN is that it reflects a prediction error response.

An important theoretical question remains whether mismatch signaling has a domain-general or domain-specific (feature-dependent) implementation in the auditory processing pathway. A recent study using invasive recordings from the cortical surface (Auksztulewicz et al., 2018) demonstrated that neural mechanisms of predictions regarding stimulus contents ("what") and timing ("when") can be dissociated in terms of their topographies and latencies throughout the frontotemporal network, and that activity in auditory regions is sensitive to interactions between different kinds of predictions. Additionally, biophysical modeling of the measured signals has shown that predictions of contents and timing are best explained either by short-term plasticity or by classical neuromodulation, respectively, suggesting separable mechanisms for signaling different kinds of predictions. However, these dissociations might be specific to predictions of contents vs. timing, which may have fundamentally different roles in processing stimulus sequences (Friston and Buzsaki, 2016).

Interestingly, an earlier magnetoencephalography (MEG) study (Phillips et al., 2015) provided evidence for a hierarchical model, whereby violations of sensory predictions regarding different stimulus contents were associated with similar response magnitudes in auditory cortex, but different connectivity patterns at hierarchically higher levels of the frontotemporal network. This result is consistent with the classical predictive coding hypothesis in which reciprocal feedforward and feedback connections at the lower levels of the hierarchy are thought to signal prediction errors and predictions regarding simple sensory features, but hierarchically higher levels are thought to signal more complex predictions and prediction errors, integrating over multiple features (Kiebel et al., 2008). Several studies, however, reported independent processing of prediction violations along different acoustic features or sound dimensions. An earlier study (Giard et al., 1995) investigated the neural correlates of mismatch processing across three different acoustic features (frequency, intensity, and duration). Mismatch responses to each feature were source-localized by fitting equivalent current dipoles to EEG signals, and the results indicated that violations of different features can be linked to dissociable sources, suggesting the involvement different underlying populations. Similar conclusions have been reached in another set of studies (Schroger, 1995; Paavilainen et al., 2001), which quantified the additivity of MMN to changes along different acoustic features, either in isolation or by combining two or more features. In these studies, the MMN response to violating two features could largely be reproduced by adding the MMN responses to violating two single features, suggesting that the latter are mutually independent. A more recent study has combined these two approaches (source localization and additivity analyses), demonstrating partial independence of three different timbre
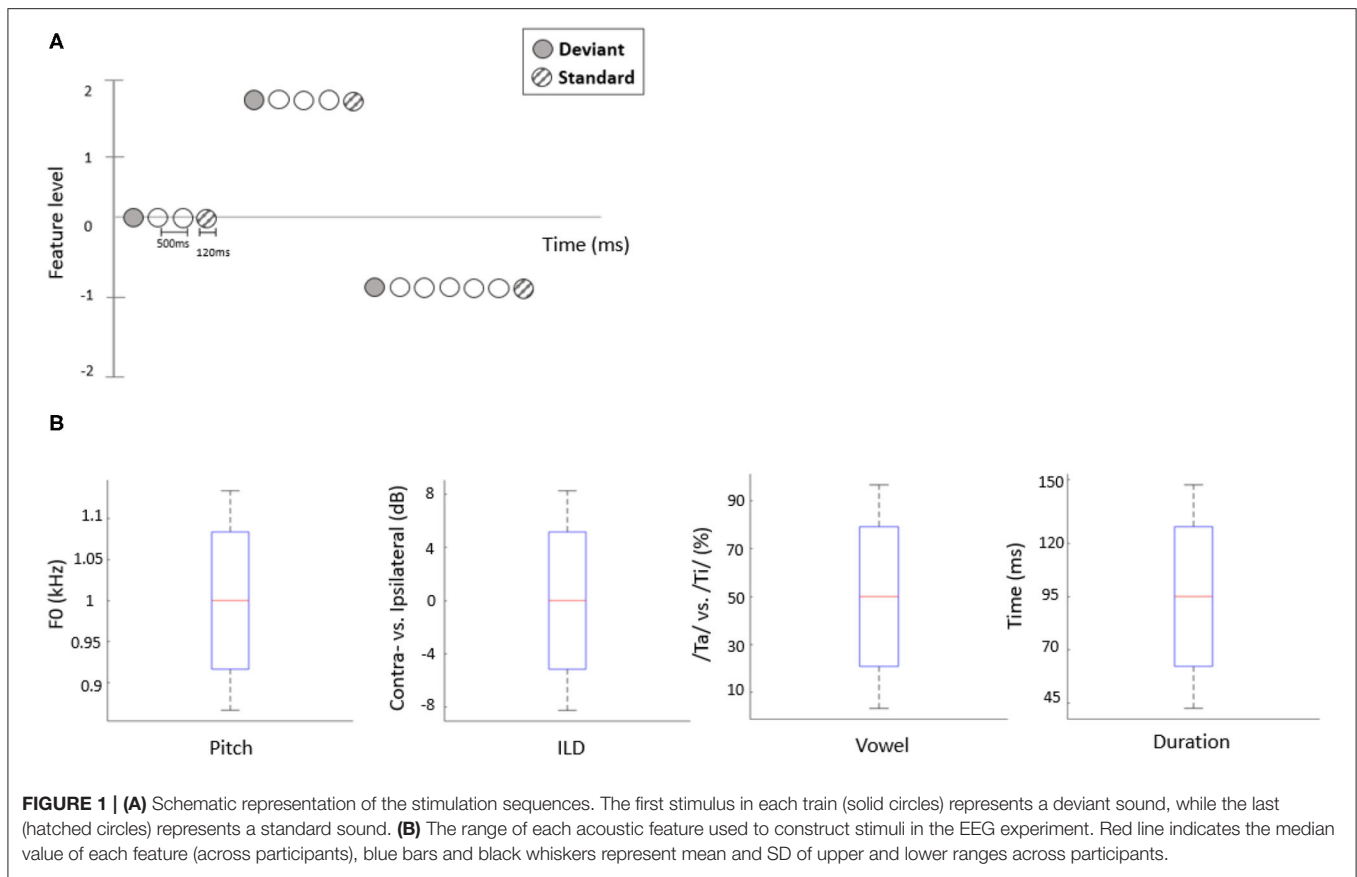
dimensions (Caclin et al., 2006). The notion that mismatch responses to violations of different features are mediated by independent mechanisms is also supported by studies showing that MMN (as well as the later P3a component) typically decreases following two identical deviants presented in direct succession, but remains stable following two deviants which vary from the standard along different features (for a review, see Rosburg et al., 2018).

However, in most previous studies (Giard et al., 1995; Schroger, 1995; Paavilainen et al., 2001; Phillips et al., 2015; Rosburg et al., 2018), physical differences between deviants and standards were not behaviorally matched across different features or participants, raising the possibility that differences in mismatch-evoked activity might to some extent be explained by differences in stimulus salience (Shiramatsu and Takahashi, 2018). This was also the case in the more recent studies on MMN responses to multiple acoustic features (Phillips et al., 2015) or in previous roving paradigms (Garrido et al., 2008). Interestingly, a recent study investigating the MMN to acoustic violations along multiple independent features in the auditory cortex of anesthetized rats (An et al., 2020) revealed that the topography of MMN signals was highly diverse across not only acoustic features but also individual animals, even though several sources of inter-subject variability (e.g., electrode placement) were better controlled than in typical non-invasive studies, suggesting that the spatial resolution of non-invasive methods such as EEG or MEG might not be sufficient for mapping more subtle differences between mismatch responses to violations of different features. The few EEG studies that did use behaviourally matched deviant sounds across different features either used very small sample sizes ($N = 8$; Deouell and Bentin, 1998) or were limited to relatively specialized perceptual characteristics (e.g., different timbre features; Caclin et al., 2006). In contrast, our study used a larger sample size ($N = 20$) and manipulated relatively general sound dimensions (location, pitch, duration, and syllable identity). Our primary goal was to test whether mismatch responses to violations of different features differ in magnitude or latency, in an attempt to replicate previous studies (Deouell and Bentin, 1998). However, in addition to testing the effects of acoustic feature on the MMN time-course in a mass-univariate analysis (i.e., on an electrode-by-electrode basis), we also aimed at decoding acoustic features from differences in MMN topography in a multivariate analysis (i.e., pooling signals from multiple electrodes).

MATERIALS AND METHODS

Participants

Twenty volunteers (13 males and 7 females; mean age 23.9 years old) enrolled in the study upon written informed consent. All participants self-reported as having normal hearing and no history of neurological disorders, and all but two were right-handed. All participants but one were native Hong Kong residents, and their mother tongue was Cantonese. A musical training questionnaire indicated that 16 participants had no musical training, and the remaining participants had <4 years' experience in playing a musical instrument. Participants were

FIGURE 1 | (A) Schematic representation of the stimulation sequences. The first stimulus in each train (solid circles) represents a deviant sound, while the last (hatched circles) represents a standard sound. **(B)** The range of each acoustic feature used to construct stimuli in the EEG experiment. Red line indicates the median value of each feature (across participants), blue bars and black whiskers represent mean and SD of upper and lower ranges across participants.

seated in a sound-attenuated and electrically shielded room in front of a computer screen. They were instructed to fixate on a fixation cross displayed on the screen during the acoustic stimulation. All experimental procedures were approved by the Human Subjects Ethics Sub-Committee of the City University of Hong Kong.

Stimuli

The present study employed a roving oddball paradigm in which auditory deviants could differ from preceding standards along one of four independent acoustic features. Specifically, we manipulated two consonant-vowel (CV) syllable stimuli, /ta/ and /ti/ (Retsa et al., 2018), along the following independent acoustic features: duration, pitch, interaural level difference (ILD) or vowel (An et al., 2020). Prior to the EEG recording, per participant, we estimated the feature interval yielding ∼80% behavioral performance by employing a 1-up-3-down staircase procedure. In each staircase trial, two out of three stimuli, chosen at random, were presented at a mean level of a given feature (e.g., a 50/50 vowel mixture or a 0 dB ILD) while the third stimulus was higher or lower than the mean level by a certain interval. Participants had to indicate which stimulus was the "odd one out." Following three consecutive hits, the interval decreased by 15%; following a mistake, the interval increased by 15%. Each participant performed 30 staircase trials for each feature (**Figure 1B**). For the roving oddball stimulus

sequences, the stimulus duration was set to 120 ms and the inter-stimulus intervals (ISIs) were fixed at 500 ms. Stimuli formed a roving oddball sequence: after 4–35 repetitions of a given stimulus (forming a standard), it was replaced with another (deviant) stimulus, randomly drawn from the set of 5 possible levels (**Figure 1A**). Roving oddball sequences corresponding to different features were administered in separate blocks, in a randomized order across participants. The total number of stimuli in each block was ∼2,000, including 200 deviant stimuli and 200 corresponding (immediately preceding) standards.

Experimental Procedure

We recorded signals from 64 EEG channels in a 10–20 system using an ANT Neuro EEG Sports amplifier. EEG channels were grounded at the nasion and referenced to the Cpz electrode. Participants were seated in a quiet room and fitted with Brainwavz B100 earphones, which delivered the audio stimuli via a MOTU Ultralite MK3 USB soundcard at 44.1 kHz. EEG signals were pre-processed using the SPM12 Toolbox for MATLAB. The continuous signals were first notch-filtered between 48 and 52 Hz and band-pass filtered between 0.1 and 90 Hz (both filters: 5th order zero-phase Butterworth), and then downsampled to 300 Hz. Eye blinks were automatically detected using the Fp1 channel, and the corresponding artifacts were removed by subtracting the two principal spatiotemporal components associated with each eye blink from all EEG channels (Ille

et al., 2002). Then, data were re-referenced to the average of all channels, segmented into epochs ranging from −100 ms before to 400 ms after each stimulus onset, baseline-corrected to the average pre-stimulus voltage, and averaged across trials to obtain ERPs for deviants and standards for each of the four acoustic features.

Data Analyses

First, to establish the presence of the MMN response, we converted the EEG time-series into 3D images (2D spatial topography × 1D time-course) and entered them into a general linear model (GLM) with two factors (random effect of mismatch: deviant vs. standard; fixed effect of participant), corresponding to a paired t-test. Statistical parametric maps were thresholded at an uncorrected $p < 0.005$, and the resulting spatiotemporal clusters of main effects were tested for statistical significance at the family-wise error corrected threshold pFWE < 0.05, taking into account the spatiotemporal correlations and multiple comparisons across channels and time points.

In an additional control analysis, we have tested whether the mismatch responses observed in this study were modulated by adaptation effects, which have been shown to be especially prominent in the N1 range (Baldeweg et al., 2004). To this end, per standard stimulus (i.e., the last stimulus in a sequence of identical stimuli), we have calculated the number of stimuli separating it from the preceding deviant (i.e., the first stimulus in a sequence of identical stimuli). If our results were indeed confounded by adaptation, the difference between responses evoked by deviants vs. standards should be modulated by the number of stimuli preceding each deviant. To test this hypothesis, we have regressed out the number of preceding stimuli from single-trial standard-evoked responses (using two regressors: a linear regressor, coding for the actual number of preceding stimuli, and a log-transformed regressor, approximating empirically observed adaptation effects; (e.g., Baldeweg et al., 2004), and subjected the residuals to the remaining univariate analysis steps (i.e., averaging the single-trial responses to obtain ERPs, and performing statistical inference while correcting for multiple comparisons across channels and time points).

Then, to test whether MMN amplitudes differed between stimulus features, ERP data were entered into a flexible-factorial GLM with one random factor (participant) and two fixed factors (mismatch: deviant vs. standard; feature: pitch, duration, ILD, and vowel), corresponding to a repeated-measures 2 × 4 ANOVA. Statistical significance thresholds were set as above.

Finally, to test whether mismatch responses can be used to decode the violated acoustic features, we subjected the data to a multivariate analysis. Prior to decoding, we calculated single-trial mismatch response signals by subtracting the EEG signal evoked by each standard from the signal evoked by the subsequent deviant. Data dimensionality was reduced using PCA (principal component analysis), resulting in spatial principal components (describing channel topographies) and temporal principal components (describing voltage time-series), sorted by the ratio of explained variance. Only those top components which, taken together, explained 95% of the original variance,
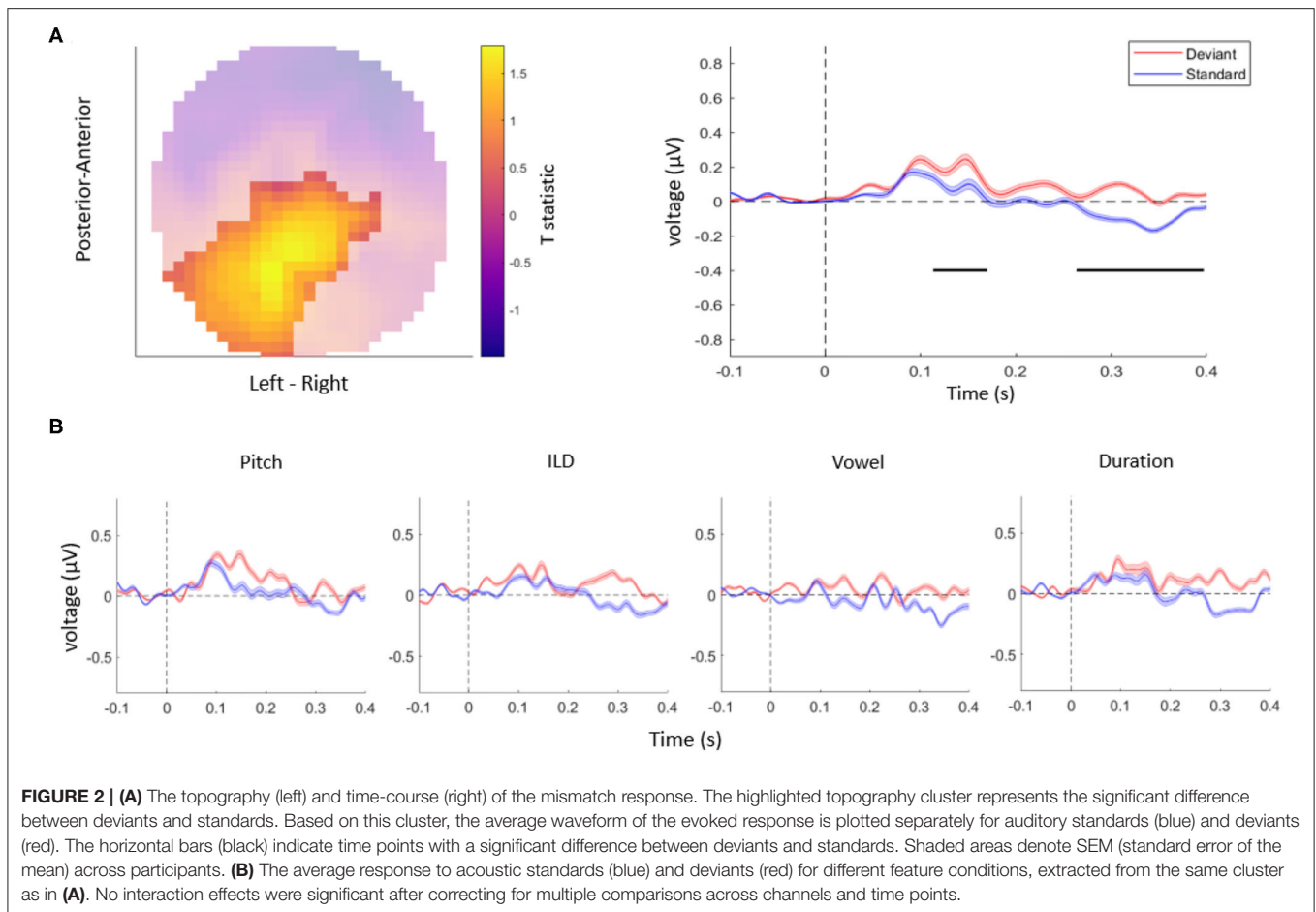
were retained for further analysis. In decoding acoustic features, we adopted a sliding window approach, integrating over the relative voltage changes within a 100 ms window around each time-point (Wolff et al., 2020). To this end, per channel and trial, the time segments within 100 ms of each analyzed time-point were down-sampled by binning the data over 10 ms bins, resulting in a vector of 10 average voltage values per component. Next, the data were de-meaned by removing the component-specific average voltage over the entire 100 ms time window from each component and time bin. These steps ensured that the multivariate analysis approach was optimized for decoding transient activation patterns (voltage fluctuations around a zero mean) at the expense of more stationary neural processes (overall differences in mean voltage) (Wolff et al., 2020).

The binned single-trial mismatch fluctuations were then concatenated across components for subsequent leave-one-out cross-validation decoding. Per trial and time point, we calculated the Mahalanobis distance (De Maesschalck et al., 2000) (scaled by the noise covariance matrix of all components) between the vector of concatenated component fluctuations of this trial (test trial) and four other vectors, obtained from the remaining trials, and corresponding to the concatenated component fluctuations averaged across trials, separately for each of the four features. The resulting Mahalanobis distance values were averaged across trials, separately for each acoustic feature, resulting in 4 × 4 distance matrices. These distance matrices were summarized per time point and participant as a single decoding estimate, by subtracting the mean off-diagonal from diagonal terms (**Figure 3A**).

In a final analysis, since we have observed univariate mismatch responses as well as multivariate mismatch-based feature decoding at similar latencies (see Results), we have tested whether these two effects are related. To this end, we performed a correlation analysis between single-trial decoding estimates (i.e., the relative Mahalanobis distance values between EEG topography corresponding to mismatch responses following violations of the same vs. different features), and single-trial MMN amplitudes. We calculated Pearson correlation coefficients across single trials, per channel, time point, and participants. The resulting correlation coefficients were subject to statistical inference using statistical parametric mapping (one-sample t-test; significance thresholds as in the other univariate analysis, corrected for multiple comparisons across time points and channels using family-wise error).

RESULTS

Taken together, in this study, we tested whether auditory mismatch responses are modulated by violations of independent acoustic features. First, consistent with previous literature (Doeller et al., 2003; Garrido et al., 2008), we observed overall differences between the ERPs evoked by deviant stimuli vs. standard stimuli, in a range typical for MMN responses as well as at longer latencies (**Figure 2A**). Specifically, the univariate ERP analysis confirmed that EEG amplitudes differed significantly between deviants and standards when pooling over

FIGURE 2 | (A) The topography (left) and time-course (right) of the mismatch response. The highlighted topography cluster represents the significant difference between deviants and standards. Based on this cluster, the average waveform of the evoked response is plotted separately for auditory standards (blue) and deviants (red). The horizontal bars (black) indicate time points with a significant difference between deviants and standards. Shaded areas denote SEM (standard error of the mean) across participants. **(B)** The average response to acoustic standards (blue) and deviants (red) for different feature conditions, extracted from the same cluster as in **(A)**. No interaction effects were significant after correcting for multiple comparisons across channels and time points.

all the acoustic features tested. This effect was observed over two clusters: the central EEG channels showed a significant mismatch response between 115 and 182 ms (cluster-level pFWE < 0.001, Tmax = 3.94), while posterior channels showed a significant mismatch response between 274 and 389 ms (cluster-level pFWE < 0.001, Tmax = 5.46), within the range of a P3b component. A control analysis, in which we controlled for single-trial adaptation effect to the standard tones, yielded a virtually identical pattern of results as the original analysis (two significant clusters of differences between responses to deviants vs. standards: an earlier cluster between 130 and 143 ms over central channels, cluster-level pFWE < 0.001, Tmax = 15.48, and a later cluster between 317 and 327 ms over posterior channels, cluster-level pFWE < 0.001, Tmax = 17.48).
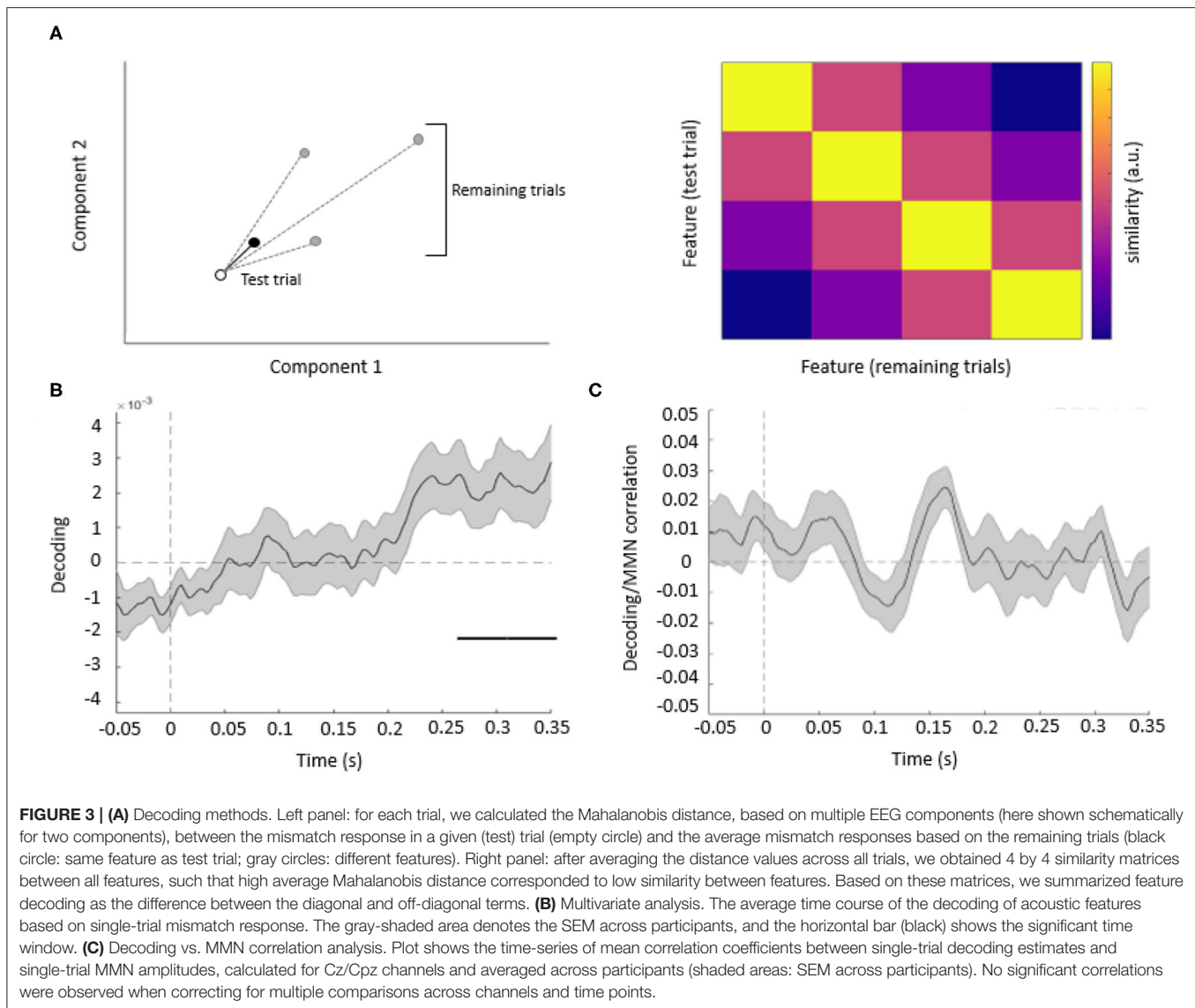
Although the ERP time-courses differed between deviant and standard stimuli when pooling over violations of different acoustic features, a univariate (channel-by-channel) analysis revealed no significant differences in the amplitudes or time-courses of mismatch responses between independent stimulus features (**Figure 2B**). These results are consistent with a previous study (Phillips et al., 2015) which found that multiple deviant stimulus features (frequency, intensity, location, duration, and silent gap) were not associated with differences in activity in the auditory regions, but instead were reflected in more distributed activity patterns (frontotemporal connectivity estimates).

The resulting decoding time-courses of each participant were entered into a GLM and subject to one-sample *t*-tests, thresholded at an uncorrected $p < 0.05$ and correcting for multiple comparisons across time points at a cluster-level pFWE < 0.05. In this analysis, significant acoustic feature decoding was observed between 247 and 350 ms relative to tone onset (cluster-level pFWE = 0.000, Tmax = 2.77) (**Figure 3B**). Thus, when pooling information from multiple EEG channels, acoustic features could be decoded from the topography of mismatch responses, although at later latencies than typical for MMN.

Since we have observed both univariate mismatch responses and multivariate mismatch-based feature decoding at late latencies (univariate: 274–389 ms; multivariate: 247–350 ms), we have performed an additional single-trial correlation analysis to test whether these two effects are related. This analysis (**Figure 3C**) has yielded no significant clusters of correlation coefficients between single-trial mismatch amplitudes and decoding estimates, while correcting for multiple comparisons across channels and time points (Tmax = 3.74, all pFWE > 0.005).

FIGURE 3 | (A) Decoding methods. Left panel: for each trial, we calculated the Mahalanobis distance, based on multiple EEG components (here shown schematically for two components), between the mismatch response in a given (test) trial (empty circle) and the average mismatch responses based on the remaining trials (black circle: same feature as test trial; gray circles: different features). Right panel: after averaging the distance values across all trials, we obtained 4 by 4 similarity matrices between all features, such that high average Mahalanobis distance corresponded to low similarity between features. Based on these matrices, we summarized feature decoding as the difference between the diagonal and off-diagonal terms. **(B)** Multivariate analysis. The average time course of the decoding of acoustic features based on single-trial mismatch response. The gray-shaded area denotes the SEM across participants, and the horizontal bar (black) shows the significant time window. **(C)** Decoding vs. MMN correlation analysis. Plot shows the time-series of mean correlation coefficients between single-trial decoding estimates and single-trial MMN amplitudes, calculated for Cz/Cpz channels and averaged across participants (shaded areas: SEM across participants). No significant correlations were observed when correcting for multiple comparisons across channels and time points.

DISCUSSION

In this study, since a univariate analysis of interactions between mismatch signals and acoustic features might not be sensitive enough to reveal subtle and distributed amplitude differences between conditions, we adopted a multivariate analysis aiming at decoding the violated acoustic feature from single-trial mismatch response topographies. This demonstrated that acoustic features could be decoded from the topography of mismatch responses, although at later latencies than typical for MMN (**Figure 3B**). An earlier oddball study (Leung et al., 2012) examined ERP differences to violations of four features (frequency, duration, intensity, and interaural difference). The study found that frequency deviants were associated with a significant amplitude change in the middle latency range. This result indicated that deviant feature detection may be subserved by a different process than general mismatch detection. Consistent with this notion, another study has used magnetoencephalography to identify mid-latency effects of local prediction violations of simple stimulus features, and contrasted them with later effects of global prediction violations of stimulus patterns (Recasens et al., 2014). Taken together, these studies would suggest that, in paradigms where multiple acoustic features vary independently (such as here), a plausible pattern of results would be that independent feature predictions should be mismatched at relatively early latencies, since an integrated representation is not required. Here, however, we found feature-specificity in the late latency range, rather than in the mid-latency range. The discrepancy between our results and the previous studies might be explained by different stimulus types. While the previous studies used simple acoustic stimuli, here we used complex syllable stimuli, possibly tapping into the later latencies of language-related mismatch responses, as compared to MMN following violations of non-speech sounds.

Speech sounds have been hypothesized to be processed in separate streams which independently derive semantic information ("what" processing) and sound location ("where" processing) (Kaas and Hackett, 2000; Tian et al., 2001; Schubotz et al., 2003; Camalier et al., 2012; Kusmierek and Rauschecker, 2014). In most animal studies, the hierarchical organization of the auditory cortex has been linked to a functional distribution of stimulus processing, such that core (hierarchically lower) regions respond preferentially to simple stimuli, whereas belt and other downstream (hierarchically higher) regions respond to more complex stimuli such as band-passed noise and speech (Rauschecker et al., 1995; Recanzone et al., 2000; Rauschecker and Tian, 2004; Kusmierek and Rauschecker, 2009; Rauschecker and Scott, 2009). This is supported by evidence functional magnetic resonance imaging (fMRI) studies in humans (Binder et al., 2000) showing that earlier auditory regions (Heschl's gyrus and surrounding fields) respond preferentially to unstructured noise stimuli, while progressively more complex stimuli such as frequency-modulated tones show more lateral response activation patterns. In that study, speech sounds showed most pronounced activations spreading ventrolaterally into the superior temporal sulcus. This result supports a hierarchical model of auditory speech processing in the human auditory cortex based on complexity and integration of temporal and spectral features. Based on this notion, the relatively long latency of neural responses compared to previous studies using pure tones might be partially explained by the fact that we used spectrally and temporally complex speech stimuli.

However, our results can also be explained in terms of a hierarchical deviance detection system based on predictive coding (Kiebel et al., 2008). On this account, neural responses supporting the lower and higher hierarchical stages communicate continuously through reciprocal pathways. When exposed to repetitive stimuli, the bottom-up (ascending) sensory inputs can be "explained away" by top-down (descending) connections mediating prediction signaling, resulting in weaker prediction error signaling back to the hierarchically higher regions. Substituting the predicted standard with unpredicted deviant results in a failure of top-down suppression by prior predictions. This leads to an increased prediction error signaling back to higher regions, providing an update for subsequent predictions. As a result, the later and more distributed activity patterns might reflect higher-order prediction errors, signaled to regions integrating multiple stimulus features and representing the entire range of stimuli likely to appear in a particular context.

In conclusion, the present study identified functional dissociations between deviance detection and deviance feature detection. First, while mismatch responses were observed at latencies typical for the MMN as well as at longer latencies, channel-by-channel analyses revealed no robust differences between mismatch responses following violations of different acoustic features. However, we demonstrate that acoustic features could be decoded at longer latencies based on fine-grained spatiotemporal patterns of mismatch responses. This finding suggests that deviance feature detection might be mediated by later and more distributed neural responses than deviance detection itself.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Human Subjects Ethics Sub-Committee of the City University of Hong Kong. The patients/participants provided their written informed consent to participate in this study. Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

AUTHOR CONTRIBUTIONS

HA: formal analysis, writing original draft, conceptualization, and conducted experiment. SH: conducted experiment and formal analysis. RA: formal analysis, supervision, project administration, and conceptualization. JS: project administration, conceptualization, investigation, and supervision. All authors contributed to the article and approved the submitted version.

FUNDING

REFERENCES

An, H., Auksztulewicz, R., Kang, H., and Schnupp, J. W. H. (2020). Cortical mapping of mismatch responses to independent acoustic features. *Hear. Res.* 399:107894. doi: 10.1016/j.heares.2020.107894

Auksztulewicz, R., Schwiedrzik, C. M., Thesen, T., Doyle, W., Devinsky, O., Nobre, A. C., et al. (2018). Not all predictions are equal: "What" and "When" predictions modulate activity in auditory cortex through different mechanisms. *J. Neurosci.* 38, 8680–8693. doi: 10.1523/JNEUROSCI.0369-18.2018

Baldeweg, T., Klugman, A., Gruzelier, J., and Hirsch, S. R. (2004). Mismatch negativity potentials and cognitive impairment in schizophrenia. *Schizophr. Res.* 69, 203–217. doi: 10.1016/j.schres.2003.09.009

Binder, J. R., Frost, J. A., Hammeke, T. A., Bellgowan, P. S., Springer, J. A., Kaufman, J. N., et al. (2000). Human temporal lobe activation by speech and nonspeech sounds. *Cereb. Cortex* 10, 512–528. doi: 10.1093/cercor/10.5.512

Caclin, A., Brattico, E., Tervaniemi, M., Naatanen, R., Morlet, D., Giard, M. H., et al. (2006). Separate neural processing of timbre dimensions in auditory sensory memory. *J. Cogn. Neurosci.* 18, 1959–1972. doi: 10.1162/jocn.2006.18.12.1959

Camalier, C. R., D'Angelo, W. R., Sterbing-D'Angelo, S. J., de la Mothe, L. A., and Hackett, T. A. (2012). Neural latencies across auditory cortex of macaque support a dorsal stream supramodal timing advantage in primates. *Proc. Natl. Acad. Sci. U.S.A.* 109, 18168–18173. doi: 10.1073/pnas.1206387109

Carbajal, G. V., and Malmierca, M. S. (2018). The neuronal basis of predictive coding along the auditory pathway: from the subcortical roots to cortical deviance detection. *Trends Hear.* 22:2331216518784822. doi: 10.1177/2331216518784822

Chennu, S., Noreika, V., Gueorguiev, D., Blenkmann, A., Kochen, S., Ibanez, A., et al. (2013). Expectation and attention in hierarchical auditory prediction. *J. Neurosci.* 33, 11194–11205. doi: 10.1523/JNEUROSCI.0114-13.2013

De Maesschalck, R., Jouan-Rimbaud, D., and Massart, D. L. (2000). The Mahalanobis distance. *Chemomet. Intellig. Lab. Syst.* 50, 1–18. doi: 10.1016/S0169-7439(99)00047-7

Deouell, L. Y., and Bentin, S. (1998). Variable cerebral responses to equally distinct deviance in four auditory dimensions: a mismatch negativity study. *Psychophysiology* 35, 745–754. doi: 10.1111/1469-8986.3560745

Doeller, C. F., Opitz, B., Mecklinger, A., Krick, C., Reith, W., and Schroger, E. (2003). Prefrontal cortex involvement in preattentive auditory deviance detection: neuroimaging and electrophysiological evidence. *Neuroimage* 20, 1270–1282. doi: 10.1016/S1053-8119(03)00389-6

Friston, K., and Buzsaki, G. (2016). The functional anatomy of time: what and when in the brain. *Trends Cogn. Sci.* 20, 500–511. doi: 10.1016/j.tics.2016.05.001

Garrido, M. I., Friston, K. J., Kiebel, S. J., Stephan, K. E., Baldeweg, T., and Kilner, J. M. (2008). The functional anatomy of the MMN: a DCM study of the roving paradigm. *Neuroimage* 42, 936–944. doi: 10.1016/j.neuroimage.2008.05.018

Giard, M. H., Lavikahen, J., Reinikainen, K., Perrin, F., Bertrand, O., Pernier, J., et al. (1995). Separate representation of stimulus frequency, intensity, and duration in auditory sensory memory: an event-related potential and dipole-model analysis. *J. Cogn. Neurosci.* 7, 133–143. doi: 10.1162/jocn.1995.7.2.133

Ille, N., Berg, P., and Scherg, M. (2002). Artifact correction of the ongoing EEG using spatial filters based on artifact and brain signal topographies. *J. Clin. Neurophysiol.* 19, 113–124. doi: 10.1097/00004691-200203000-00002

Kaas, J. H., and Hackett, T. A. (2000). Subdivisions of auditory cortex and processing streams in primates. *Proc. Natl. Acad. Sci. U.S.A.* 97, 11793–11799. doi: 10.1073/pnas.97.22.11793

Kiebel, S. J., Daunizeau, J., and Friston, K. J. (2008). A hierarchy of time-scales and the brain. *PLoS Comput. Biol.* 4:e1000209. doi: 10.1371/journal.pcbi.1000209

Kusmierek, P., and Rauschecker, J. P. (2009). Functional specialization of medial auditory belt cortex in the alert rhesus monkey. *J. Neurophysiol.* 102, 1606–1622. doi: 10.1152/jn.00167.2009

Kusmierek, P., and Rauschecker, J. P. (2014). Selectivity for space and time in early areas of the auditory dorsal stream in the rhesus monkey. *J. Neurophysiol.* 111, 1671–1685. doi: 10.1152/jn.00436.2013

Leung, S., Cornella, M., Grimm, S., and Escera, C. (2012). Is fast auditory change detection feature specific? An electrophysiological study in humans. *Psychophysiology* 49, 933–942. doi: 10.1111/j.1469-8986.2012.01375.x

Naatanen, R. (2007). The mismatch negativity - where is the big fish? *J. Psychophysiol.* 21, 133–137. doi: 10.1027/0269-8803.21.34.133

Paavilainen, P., Valppu, S., and Naatanen, R. (2001). The additivity of the auditory feature analysis in the human brain as indexed by the mismatch negativity: 1+1 approximate to 2 but 1+1+1 < 3. *Neurosci. Lett.* 301, 179–182. doi: 10.1016/S0304-3940(01)01635-4

Phillips, H. N., Blenkmann, A., Hughes, L. E., Bekinschtein, T. A., and Rowe, J. B. (2015). Hierarchical organization of frontotemporal networks for the prediction of stimuli across multiple dimensions. *J. Neurosci.* 35, 9255–9264. doi: 10.1523/JNEUROSCI.5095-14.2015

Rauschecker, J. P., and Scott, S. K. (2009). Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. *Nat. Neurosci.* 12, 718–724. doi: 10.1038/nn.2331

Rauschecker, J. P., and Tian, B. (2004). Processing of band-passed noise in the lateral auditory belt cortex of the rhesus monkey. *J. Neurophysiol.* 91, 2578–2589. doi: 10.1152/jn.00834.2003

Rauschecker, J. P., Tian, B., and Hauser, M. (1995). Processing of complex sounds in the macaque nonprimary auditory cortex. *Science* 268, 111–114. doi: 10.1126/science.7701330

Recanzone, G. H., Guard, D. C., and Phan, M. L. (2000). Frequency and intensity response properties of single neurons in the auditory cortex of the behaving macaque monkey. *J. Neurophysiol.* 83, 2315–2331. doi: 10.1152/jn.2000.83.4.2315

Recasens, M., Grimm, S., Wollbrink, A., Pantev, C., and Escera, C. (2014). Encoding of nested levels of acoustic regularity in hierarchically organized areas of the human auditory cortex. *Hum. Brain Mapp.* 35, 5701–5716. doi: 10.1002/hbm.22582

Retsa, C., Matusz, P. J., Schnupp, J. W. H., and Murray, M. M. (2018). What's what in auditory cortices? *Neuroimage* 176, 29–40. doi: 10.1016/j.neuroimage.2018.04.028

Rosburg, T., Weigl, M., Thiel, R., and Mager, R. (2018). The event-related potential component P3a is diminished by identical deviance repetition, but not by non-identical repetitions. *Exp. Brain Res.* 236, 1519–1530. doi: 10.1007/s00221-018-5237-z

Schroger, E. (1995). Processing of auditory deviants with changes in one versus two stimulus dimensions. *Psychophysiology* 32, 55–65. doi: 10.1111/j.1469-8986.1995.tb03406.x

Schubotz, R. I., von Cramon, D. Y., and Lohmann, G. (2003). Auditory what, where, and when: a sensory somatotopy in lateral premotor cortex. *Neuroimage* 20, 173–185. doi: 10.1016/S1053-8119(03)00218-0

Shiramatsu, T. I., and Takahashi, H. (2018). Mismatch negativity in rat auditory cortex represents the empirical salience of sounds. *Front. Neurosci.* 12:924. doi: 10.3389/fnins.2018.00924

Tian, B., Reser, D., Durham, A., Kustov, A., and Rauschecker, J. P. (2001). Functional specialization in rhesus monkey auditory cortex. *Science* 292, 290–293. doi: 10.1126/science.1058911

Wolff, M. J., Kandemir, G., Stokes, M. G., and Akyurek, E. G. (2020). Unimodal and bimodal access to sensory working memories by auditory and visual impulses. *J. Neurosci.* 40, 671–681. doi: 10.1523/JNEUROSCI.1194-19.2019