

Prosodic pitch processing is represented in delta-band EEG and is dissociable from the cortical tracking of other acoustic and phonetic features

Emily S. Teoh¹  | Madeline S. Cappelloni² | Edmund C. Lalor^{1,2,3}

¹School of Engineering, Trinity Centre for Biomedical Engineering and Trinity College Institute of Neuroscience, Trinity College Dublin, University of Dublin, Dublin, Ireland

²Department of Biomedical Engineering, University of Rochester, Rochester, NY, USA

³Department of Neuroscience and Del Monte Institute for Neuroscience, University of Rochester, Rochester, NY, USA

Correspondence

Emily S. Teoh, School of Engineering, Trinity Centre for Biomedical Engineering and Trinity College Institute of Neuroscience, Trinity College Dublin, Dublin 2, Ireland.
Email: teohe@tcd.ie

Funding information

Irish Research Council, Grant/Award Number: GOIPG/2015/3378

The peer review history for this article is available at <https://publons.com/publon/10.1111/EJN.14510>

Abstract

Speech is central to communication among humans. Meaning is largely conveyed by the selection of linguistic units such as words, phrases and sentences. However, prosody, that is the variation of acoustic cues that tie linguistic segments together, adds another layer of meaning. There are various features underlying prosody, one of the most important being pitch and how it is modulated. Recent fMRI and ECoG studies have suggested that there are cortical regions for pitch which respond primarily to resolved harmonics and that high-gamma cortical activity encodes intonation as represented by relative pitch. Importantly, this latter result was shown to be independent of the cortical tracking of the acoustic energy of speech, a commonly used measure. Here, we investigate whether we can isolate low-frequency EEG indices of pitch processing of continuous narrative speech from those reflecting the tracking of other acoustic and phonetic features. Harmonic resolvability was found to contain unique predictive power in delta and theta phase, but it was highly correlated with the envelope and tracked even when stimuli were pitch-impovertished. As such, we are circumspect about whether its contribution is truly pitch-specific. Crucially however, we found a unique contribution of relative pitch to EEG delta-phase prediction, and this tracking was absent when subjects listened to pitch-impovertished stimuli. This finding suggests the possibility of a separate processing stream for prosody that might operate in parallel to acoustic-linguistic processing. Furthermore, it provides a novel neural index that could be useful for testing prosodic encoding in populations with speech processing deficits and for improving cognitively controlled hearing aids.

KEYWORDS

EEG, pitch, prosody

Abbreviations: (m)TRF, (multivariate) temporal response function; ECoG, electrocorticography; EEG, electroencephalography; *env*, speech envelope; fMRI, functional magnetic resonance imaging; f_{rel} , relative pitch; MEG, magnetoencephalography; *res*, resolvability.

Edited by Gregor Thut.

1 | INTRODUCTION

Speech is arguably humanity's most important signal because of its extraordinary power to convey information. It is built up by a hierarchy of linguistic units—syllables construct words, which in themselves have meaning, and words are organized into phrases to express more complex ideas. But when we speak, these different segments are not simply uttered at a fixed rate and tone. We thread them together with rhythms and melodies that inform how the speech is to be parsed (Steinhauer, Alter, & Friederici, 1999) and communicate other layers of meaning. These suprasegmental variations of acoustic cues—the lengthening of a syllable, the rise and fall of pitch across a phrase, the increase in loudness at a word—make up speech prosody. Indeed, the same sequence of words can convey very different meanings and emotions by virtue of prosodic differences (Bolinger & Bolinger, 1986).

Because of prosody's importance, it is likely that we have evolved specific neurophysiological mechanisms to process this additional channel of information. Many studies have sought to anatomically localize prosodic processing in the brain (Baum & Pell, 1999; Gandour et al., 2004; Kreitewolf, Friederici, & von Kriegstein, 2014; Meyer, Alter, Friederici, Lohmann, & von Cramon, 2002; Plante, Creusere, & Sabin, 2002; Sammler, Grosbras, Anwander, Bestelmeyer, & Belin, 2015; Witteman, van Ijzendoorn, van de Velde, van Heuven, & Schiller, 2011). But a comprehensive model of how the brain processes speech necessitates understanding precisely how prosodic information is encoded in the brain. The quest to uncover its neural correlates is complicated by the fact that a prosodic percept can arise from the modulation of one or multiple acoustic measures, and it remains unclear to what extent individual cues or cue combinations contribute to a particular prosodic feature (Cole & Shattuck-Hufnagel, 2016). That said, an undeniably important auditory percept, the modulation of which underlies various prosodic phenomena, is pitch, which can be estimated from the acoustic speech stream (Boersma, 1993).

It has been established that there are auditory cortical regions that encode sound pitch (Briley, Breakey, & Krumbholz, 2013; Griffiths et al., 2010; Hall & Plack, 2009) and that these pitch-sensitive regions respond primarily to resolved harmonics (Norman-Haignere, Kanwisher, & McDermott, 2013). But in the context of speech, evidence has only recently been found for the cortical encoding of speaker-normalized relative pitch. Tang, Hamilton, and Chang (2017) measured high-gamma electrocorticography (ECoG) responses to set of spoken sentences that independently varied intonation contour, phonetic content and speaker identity, and found high-gamma activity on single electrodes over superior temporal gyrus selectively represented intonation contours in terms of relative pitch. This demonstrates how intonation

undergoes specialized extraction from the speech signal, distinct from other aspects of speech.

In the present study, we investigate the encoding of pitch during listening to continuous narrative speech in ongoing scalp-recorded electroencephalography (EEG). Like ECoG, EEG has high temporal resolution but with markedly less sensitivity to high-gamma activity. However, EEG has the advantages of being portable and non-invasive, meaning that it can be used to study speech and language processing in a wider variety of subjects. It is well-established that low-frequency EEG tracks the temporal speech envelope (Aiken & Picton, 2008; Ding & Simon, 2014), and it was recently shown that using the spectrogram to represent acoustic energy leads to improved EEG prediction (Di Liberto, O'Sullivan, & Lalor, 2015). But disambiguating responses to other features related to the neural processing of speech can be difficult as most features covary and are nested in the envelope. Isolable neural indices at the level of phonemic (Di Liberto et al., 2015) and semantic processing (Broderick, Anderson, Liberto, Crosse, & Lalor, 2018) have recently been found, but it is unknown whether pitch processing can also be dissociated.

We employ an established approach (Lalor & Foxe, 2010) to regress EEG on two pitch-related measures—harmonic resolvability (Norman-Haignere et al., 2013), on which pitch saliency is dependent, and relative pitch (Tang et al., 2017). We test whether these features significantly predict EEG after accounting for features previously shown to index acoustic/phonetic processing and examine their response characteristics. Given that different EEG frequency bands have been linked to different roles in speech processing (Giraud & Poeppel, 2012), we also examine whether these indices may be more evident in particular bands.

2 | MATERIALS AND METHODS

2.1 | Subjects

All subjects spoke English as their primary language and had no reported history of hearing impairment or neurological disorder. Our study involved analysing data from two experiments. Nineteen subjects (13 male) took part in the first experiment and 13 subjects (eight male) took part in the second. Both experiments were undertaken in accordance with the Declaration of Helsinki and were approved by the Ethics Committee of the School of Psychology at Trinity College Dublin. Each subject provided written informed consent. These data have been previously analysed using different methods and published (experiment 1—Broderick et al., 2018; Di Liberto et al., 2015; experiment 2—Di Liberto, Crosse, & Lalor, 2018). Data for the first experiment can be found at <http://datadryad.org/resource/doi:10.5061/dryad.070jc>.

2.2 | Experimental design

In the first experiment, subjects undertook 20 trials, each of between 2 and 3 min in length, in which they were presented with an audiobook version of a classic work of fiction (“The Old Man and the Sea” by Ernest Hemmingway) read by a professional male American English narrator. Subjects were asked to simply listen to the story (there was no behavioural task), and they were allowed to take breaks in between trials.

The second experiment was originally designed to test the effects of prior knowledge on the encoding of natural speech (Di Liberto et al., 2018). We utilized data from a segment of the experiment to validate our pitch-related measures. In this particular segment, subjects were presented with a 10-s snippet of noise-vocoded (Davis & Johnsrude, 2003; Shannon, Zeng, Kamath, Wygonski, & Ekelid, 1995) speech, followed by a clean speech snippet. The noise-vocoded and clean speech snippets have highly correlated acoustic envelopes, but whilst clean speech possesses clear pitch, vocoded speech does not. The snippets were randomly selected from the same audiobook as the first experiment. In the majority of trials (“standard” trials), the noise-vocoded speech snippet was a degraded version of the subsequently presented clean speech snippet (specifically, the clean speech was filtered into three frequency bands and the amplitude envelope of each band was used to modulate band-limited noise). However, in some of the trials, the vocoded and clean speech snippets did not match (“deviant” trials). Subjects were tasked with identifying if a trial was standard or deviant after listening to the clean speech. We only analysed data from standard trials (78 ten-second trials in total per subject).

All stimuli were presented monophonically at a sampling rate of 44,100 Hz using Sennheiser HD650 headphones and Presentation software from Neurobehavioral Systems (<http://www.neurobs.com>). Testing was carried out in a dark room, and subjects were instructed to maintain visual fixation for the duration of each trial on a crosshair centred on the screen and to minimize eye blinking and all other motor activities.

2.3 | Data acquisition and preprocessing

A BioSemi ActiveTwo system was used to record EEG data from 128 electrode positions digitized at 512 Hz in both experiments. All channels were referenced to the average of two mastoid channels and downsampled to 128 Hz. EEG data were filtered between 0.2 and 30 Hz (broadband), and into delta (0.2–4 Hz), theta (4–8 Hz), alpha (8–15 Hz) and beta (15–30 Hz) bands using Butterworth filters. Instantaneous power and phase were computed for the narrowband filtered data by taking the absolute value and the cosine of the angle of the Hilbert transform of the signals, respectively.

We represent our speech stimuli in terms of two pitch-related parameters—pitch (f_{rel}) and harmonic resolvability (res) (example waveforms of the pitch measures, as well as the

temporal envelope for comparison, are shown in Figure 1). f_{rel} quantifies pitch, normalized according to the vocal range of the speaker, and res is a measure related to whether the harmonics of a sound can be processed within distinct filters of the cochlea (resolved) or if they interact within the same filter (unresolved). It was important to ensure that any neural tracking of these pitch measures was not explainable in terms of the various acoustic and phonetic features that have previously been shown to predict neural responses (Di Liberto et al., 2015). As such, we included a number of other speech representations in our analysis framework, including acoustic energy measures—the temporal envelope (env) and spectrogram (spec)—as well as phonetic features (fea). The measures were computed as follows:

Pitch-related measures -

- f_{rel} : Praat software (Boersma & Weenink, 2016) was used to extract a continuous measure of pitch (fundamental frequency/absolute pitch) at a sampling rate of 128 Hz. It performs pitch estimation using an autocorrelation method (Boersma, 1993). The z -score of this measure was then computed to obtain speaker-normalized relative pitch.

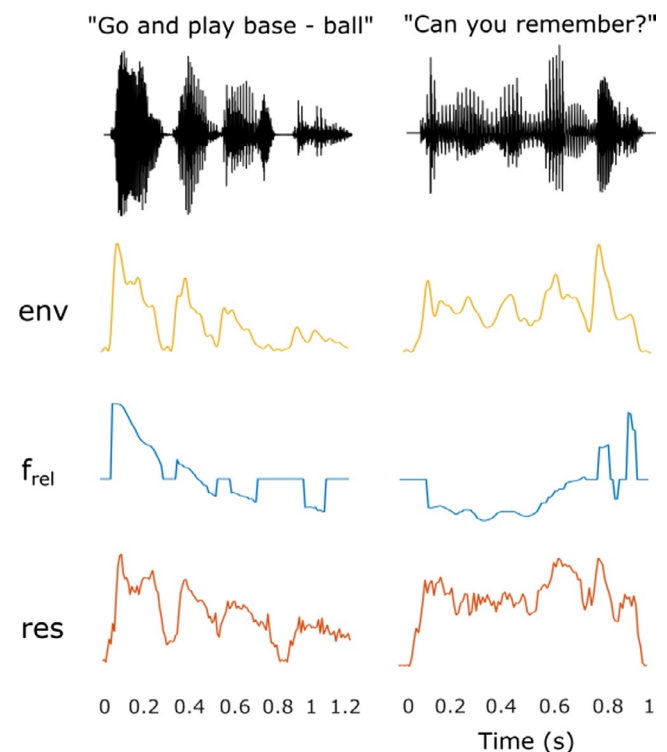


FIGURE 1 Subjects listened to an audiobook version of a classic work of fiction read by a professional narrator. Pitch-related measures—relative pitch (f_{rel}) and harmonic resolvability (res)—and the temporal envelope (env) estimated from the audio signal of example statement (“Go and play baseball”) and question (“Can you remember?”) phrases from the book are depicted. Statements in English generally end with a falling intonation contour, whilst questions end with a rising contour as can be seen in the f_{rel} measure. [Colour figure can be viewed at wileyonlinelibrary.com]

- *res*: Following fMRI work on this topic (Norman-Haignere et al., 2013), we extracted a measure of harmonic resolvability based on a model of the human auditory periphery (McDermott & Simoncelli, 2011).

Acoustic/phonetic representations (collectively, *efs*) -

- *env*: The stimuli were first filtered between 80 and 2,800 Hz. The absolute value of the Hilbert transform of the resulting signal was then taken. The envelopes were further low-pass filtered at 30 Hz and downsampled to 128 Hz.
- *spec*: Stimuli were filtered into 16 frequency bands between 250 and 8 kHz according to Greenwood's equation (Greenwood, 1961), and then, the amplitude envelope for each frequency band was computed by taking the absolute value of the Hilbert transform.
- *fea*: Prosodylab-Aligner software (Gorman, Howell, & Wagner, 2011) was used to perform forced-alignment, generating phoneme-level (American English International Phonetic Alphabet; IPA) segmentation of the stimuli. Each phoneme was then mapped to a set of 19 features based on their manner of articulation (University of Iowa's phonetics project; <http://prosodylab.cs.mcgill.ca/tools/aligner/>). Based on the start and end times of each feature, a multivariate time-series binary matrix was produced to mark when each feature was present in the stimuli.

The Pearson correlation between f_{rel} and *res* with the envelope was found to be .187 and .843, respectively.

2.4 | Multivariate linear regression

The different speech representations described above were mapped to the concurrently recorded 128-channel EEG signals. Multivariate regularized linear regression was employed to relate the speech features to the recorded EEG data (forward modelling), where each EEG channel is estimated to be a linear transformation of the speech features over a range of time lags. This transformation is described by the so-called temporal response function (TRF). For a particular speech feature, this operation can be represented mathematically as following (Crosse, Di Liberto, Bednar, & Lalor, 2016):

$$r(t, n) = \sum_{\tau} w(\tau, n) s(t - \tau) + \epsilon(t, n)$$

where $r(t, n)$ is the neural (EEG) response at channel, n , and time point, $t = 1 \dots T$, $s(t - \tau)$, is the stimulus at a lag, τ , $w(\tau, n)$ is the transformation (TRF) of the stimulus at lag τ , and $\epsilon(t, n)$ is the residual response not explained by the model.

The TRF is estimated by minimizing the mean square error between the actual neural response, $r(t, n)$, and the response predicted by the transformation, $\hat{r}(t, n)$. In practice, this can be solved by using reverse correlation. We use the mTRF toolbox (Crosse et al., 2016; <https://sourceforge.net/projects/aespa/>), which solves for the TRF (w) using reverse correlation with ridge regression:

$$w = (S^T S + \lambda I)^{-1} S^T r$$

where λ is the ridge regression parameter, I is the identity matrix, and the matrix S is the lagged time series of the stimulus matrix, s . The TRF approach can be used to relate multiple features of the stimuli to the ongoing EEG simultaneously by extending the lagged stimulus matrix to include the various features (more details can be found in Crosse et al., 2016). The ridge regression parameter was tuned using leave-one-out cross-validation. That is, we trained on $n-1$ trials for a wide range of λ values (e.g. 1×101 , 1×102 , ... 1×109), computed the average TRF across trials for each λ , then tested the TRFs on the n th trial. This was repeated n times, rotating the trial to be tested each time. The λ value that maximized the Pearson correlation coefficient between the actual and predicted neural response over all trials was selected.

The transformation was computed over time lag intervals of 0–300 ms. This was selected based on previous speech-related studies, where no visible response was present outside this range (Di Liberto et al., 2015; Lalor & Foxe, 2010). Using a time lag interval reflects the idea that changes in the features of the ongoing stimulus are likely to produce effects in the ongoing EEG in that interval. We quantified how well each speech representation related to the neural data using leave-one-trial-out cross-validation (as described above), with the Pearson correlation coefficient as our metric of prediction accuracy. Because the cross-validation procedure takes the average of the validation metric across trials, the models are not biased towards the test data used for cross-validation (Crosse et al., 2016).

To evaluate whether a pitch feature contributed independently of all other features in predicting the neural data, we computed the partial correlation coefficients (Pearson's r) between the EEG predicted by each pitch measure's model with the actual recorded EEG after controlling for the effects of all other acoustic/phonetic and pitch features that were found to predict EEG at above chance level.

2.5 | Statistical analysis

To test that an individual feature predicts EEG at above chance level, we performed non-parametric permutation testing. The neural responses were permuted across trials such that they were matched to features from a different trial, and the same

leave-one-out cross-validation procedure as described above was performed to compute TRFs and prediction accuracies. This was done 1,000 times for each subject to establish a distribution of chance-level prediction accuracies. Based on these prediction accuracies, we also computed a distribution of partial correlations for the pitch-related measures (i.e. we partialled out the contributions of all other features).

To perform group-level statistical testing, we generated a null distribution of group means. One prediction accuracy from each subject's individual distribution was selected at random to go into each group mean. This process was repeated 1,000 times, sampling with replacement for each subject.

3 | RESULTS

3.1 | Pitch-related features uniquely contribute to predictions of broadband EEG

We regressed the individual pitch and acoustic/phonetic representations to broadband EEG (0.2–30 Hz). Figure 2 shows the prediction accuracies for the pitch-related measures (averaged over the best 30 channels for each feature; but using all channels did not alter the pattern of results), along with their topographic distributions. We found that all features could individually predict EEG better than chance (one-tailed permutation test, f_{rel} : $p = 9.99\text{e-}4$, res :

$p = 9.99\text{e-}4$, env : $p = 9.99\text{e-}4$, fea : $p = 9.99\text{e-}4$, $spec$: $p = 9.99\text{e-}4$). Of course, there is some redundancy in these features—there is overlap between the acoustic/phonetic measures, and as mentioned above, we found correlation between the pitch features and envelope. Given that the acoustic/phonetic features are tracked by cortex, it is therefore not surprising to also find significant tracking of the two pitch-related measures. One way to examine whether they contribute any unique predictive power is to test for improvement in prediction accuracy when a pitch-related feature is added to the joint acoustic/phonetic model ($env + fea + spec$, or efs for short). We found this to be the case—both pitch-related measures improved EEG prediction above and beyond that of the joint acoustic/phonetic model (Wilcoxon signed-rank test [vs. efs] – $f_{\text{rel}} + efs$: $z = 3.299$, $p = 9.6733\text{e-}04$, $res + efs$: $z = 2.213$, $p = .0269$, $f_{\text{rel}} + res + efs$: $z = 3.702$, $p = 2.1367\text{e-}04$). The results of these combined models are also depicted in Figure 2.

A second way of checking for unique predictive power is to compute the partial correlations of the EEG predicted by the pitch-related features with the actual EEG after controlling for all other features (including the other pitch representation). An advantage of this method is that the specific contribution of each measure can be more clearly characterized. Figure 3 depicts the partial correlations of the pitch features and their topographic distributions. Non-parametric permutation testing revealed significant

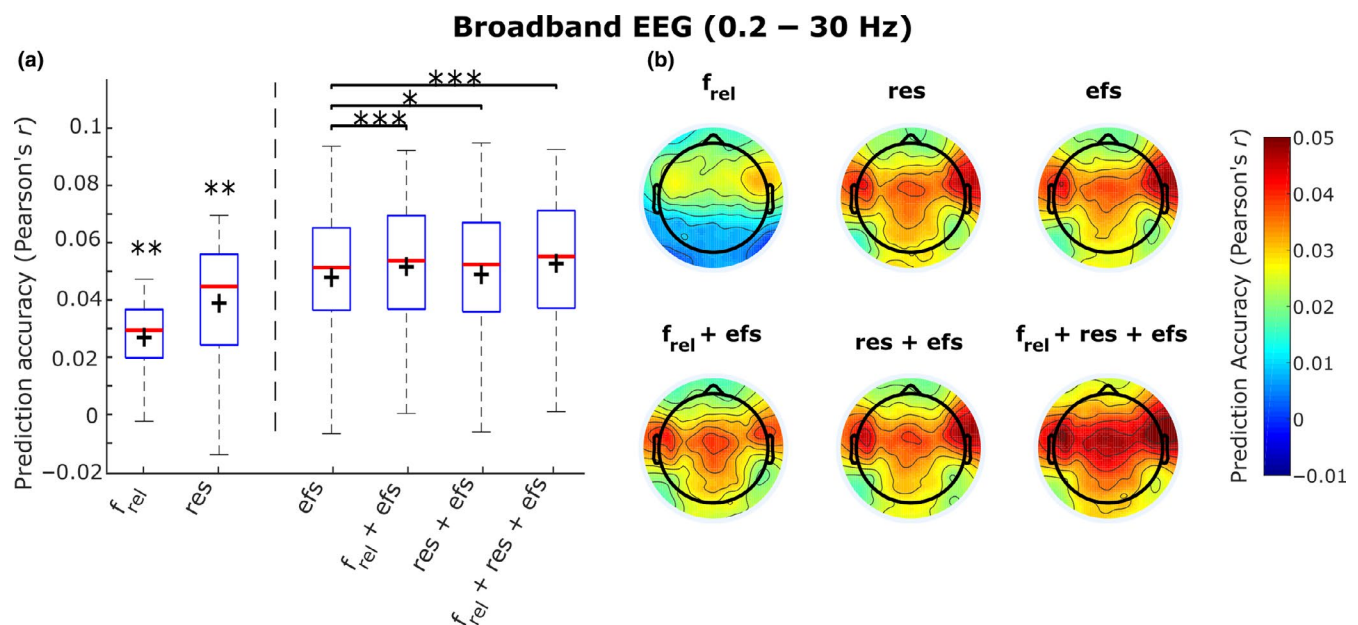


FIGURE 2 Broadband EEG prediction accuracies (Pearson's r) for the two pitch-related measures (f_{rel} & res ; averaged across the best 30 channels for each representation/combination, although using all channels did not alter the significance of results) are shown on the left side of plot (a). To determine whether these features have unique predictive power, we check whether their inclusion improves prediction accuracies above and beyond the joint acoustic/phonetic model (efs). The prediction accuracies of these combined models are shown on the right side of plot (a). The black crosses indicate mean across all subjects, and asterisks indicate statistical significance (for individual measures: one-tailed permutation test; for paired comparison of the combined models to efs : Wilcoxon signed-rank test; $*p < .05$, $**p < .01$, $***p < .001$). Topographic distribution of prediction accuracies across all subjects is shown in (b). [Colour figure can be viewed at wileyonlinelibrary.com]

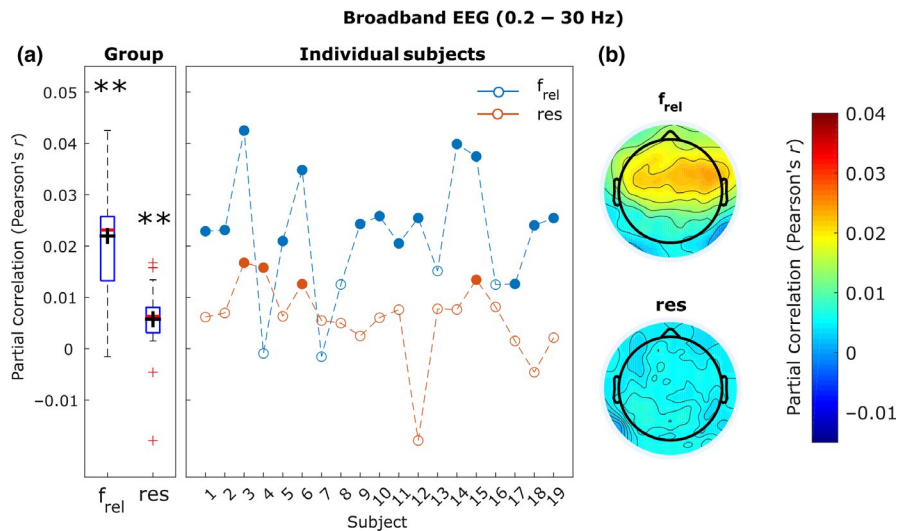


FIGURE 3 Partial correlations of broadband EEG predicted by the two pitch-related measures, f_{rel} and res , after accounting for all acoustic and phonetic measures (average over 30 best channels). (a) Group- and individual-level results; the black crosses in the group plot indicate the mean across all subjects for each feature and asterisks indicate statistical significance (one-tailed permutation test, $*p < .05$, $**p < .01$). For individual subjects, filled circles indicate statistical significance above chance (one-tailed permutation test, $p < .05$). (b) Topographic distribution of the partial correlations for pitch-related measures across all subjects. [Colour figure can be viewed at wileyonlinelibrary.com]

partial correlations (on a group level) for both measures (one-tailed permutation test, f_{rel} : $p = 9.99\text{e-}4$, and res : $p = .0020$). On an individual subject level, we found significant partial correlation for 14 of the 19 subjects for f_{rel} , and 4 of the 19 subjects for res .

3.2 | Relative pitch tracking is specific to delta phase, but all other features are indexed in both delta and theta phases

To more precisely identify which components of the EEG signal best encode pitch-related features, we computed the power and phase of delta-, theta-, alpha- and beta-band EEG. We repeated the aforementioned regression analysis, mapping our stimulus features to each of these EEG components. We found that, consistent with speech perception literature, our measures were better reflected in the delta- and theta-phase components than in all other analytic components, with f_{rel} primarily reflected in delta phase. We therefore focused on the delta- and theta-phase components.

In the delta band, all features could individually predict EEG phase better than chance on a group level (one-tailed permutation test, env : $p = 9.99\text{e-}4$, $spec$: $p = 9.99\text{e-}4$, fea : $p = 9.99\text{e-}4$, f_{rel} : $p = 9.99\text{e-}4$, and res : $p = 9.99\text{e-}4$). As before, we computed the partial correlations of the two pitch-related measures. Both features had significant partial correlations (f_{rel} : $p = 9.99\text{e-}4$, and res : $p = 9.99\text{e-}4$), suggesting that they contributed uniquely to predicting delta phase beyond the other features. On an individual subject level, we found significant partial correlation values in 16 of the 19 subjects for f_{rel} and 5 of the 19 subjects for res . The group and individual partial correlations of the pitch-related measures are shown in Figure 4a.

In the theta band, again all features could individually predict EEG phase better than chance on a group level (one-tailed

permutation test, env : $p = 9.99\text{e-}4$, $spec$: $p = 9.99\text{e-}4$, fea : $p = 9.99\text{e-}4$, f_{rel} : $p = 9.99\text{e-}4$, and res : $p = 9.99\text{e-}4$). However, when we partialled out the contribution of all other features from the EEG predicted by the pitch-related measures, we found that of the two measures, only res had a unique contribution to predicting theta phase ($p = 9.99\text{e-}4$)—the partial correlation of f_{rel} was not significantly greater than chance ($p = .076$). On an individual subject level, we found significant partial correlation values in 3 of the 19 subjects for f_{rel} and 13 of the 19 subjects for res . Partial correlations for theta phase are shown in Figure 4b.

We examined the topographic distributions of the partial correlations on the scalp (i.e. how well data on different channels could be predicted) for delta and theta phase (Figure 4c,d). Based on the topographic plots, f_{rel} appears to be more strongly encoded in mid-frontal channels in delta-phase EEG, but there were no visible clusters with higher partial correlations for f_{rel} in the theta band. A different pattern of results was observed for res —there were no visible clusters with higher partial correlations in delta-phase EEG, but two clusters (one on each hemisphere) were noticeable in the theta band.

As well as examining how well data on each channel can be predicted by TRF models, one can also visualize the TRF model weights themselves with a view to understanding what channels show a relationship to the various speech features and at what time lags those relationships occur (Crosse et al., 2016). This is because the TRF reflects how the instantaneous value of a stimulus feature impacts upon the response. We examined the TRFs of the pitch-related features, focusing again on the phase of delta and theta EEG (Figure 5). To control for the effect of all other features, we computed a joint TRF model using all acoustic/phonetic and pitch features and extracted only weights corresponding to our pitch-related measures from the model.

FIGURE 4 Group and single-subject level partial correlations (Pearson's r ; average over 30 best channels) of pitch-related features when predicting the phase for the (a) delta and (b) theta frequency bands of the EEG signal, after controlling for all other measures. Asterisks in the group plots indicate statistical significance (one-tailed permutation test, $*p < .05$, $**p < .01$), and black crosses indicate the mean across all subjects. For individual subjects, filled circles indicate statistical significance above chance (one-tailed permutation test, $p < .05$). Also shown are topographic distributions (c & d) of the partial correlations to delta and theta phase, respectively. [Colour figure can be viewed at wileyonlinelibrary.com]

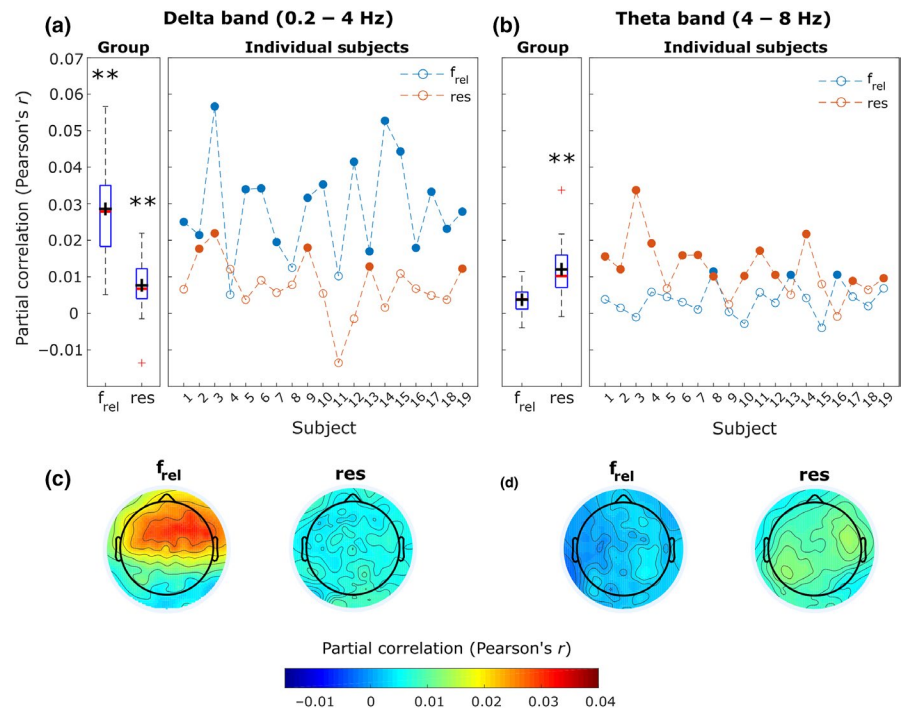
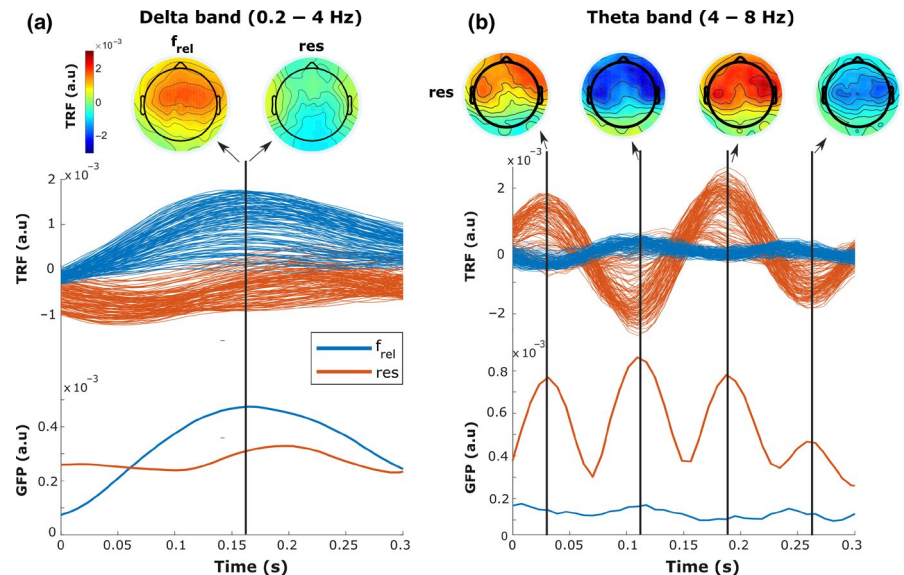


FIGURE 5 TRF weights of all channels (averaged over subjects) that reflect the mapping of individual pitch-related features (obtained by fitting a joint model to all features and then extracting weights corresponding to pitch-related measures) to (a) delta-phase and (b) theta-phase EEG. Topographic distributions of the weights are shown at ~ 0.16 s for f_{rel} and res in delta phase, and at ~ 0.03 s, ~ 0.11 s, ~ 0.19 s and ~ 0.26 s for res in theta phase (not shown for f_{rel} as it does not contribute uniquely above chance in theta). These time points correspond to the peaks in the global field power (GFP)—shown below the TRFs. [Colour figure can be viewed at wileyonlinelibrary.com]



For delta phase, the TRF of f_{rel} was a monophasic signal which peaked in power at ~ 160 ms. The TRF of res had a smaller magnitude and peaked later (~ 200 ms) but for comparison, we show the topographic distribution of TRF weights for both measures at ~ 160 ms in Figure 5a. The two measures display distinct topographies, with fronto-central positivity for f_{rel} and bilateral fronto-temporal positivity and posterior negativity for res . For theta phase, the weights for f_{rel} were small in magnitude; as this measure does not significantly predict theta phase, we did not plot the topography of its weights. The TRF for res was a quadriphasic signal that peaked at ~ 30 , ~ 110 , ~ 190 and ~ 260 ms. At these times, the

weights were highest in magnitude at bilateral fronto-temporal regions.

3.3 | Noise-vocoded speech: Validation that f_{rel} indexes pitch-specific activity

As a check that our pitch-related measures reflect pitch-specific activity, we regressed acoustic and pitch-related measures extracted from clean speech to EEG data (delta and theta phase) recorded during experiment 2. As mentioned above, this experiment involved participants listening to both clean and noise-vocoded versions of the same

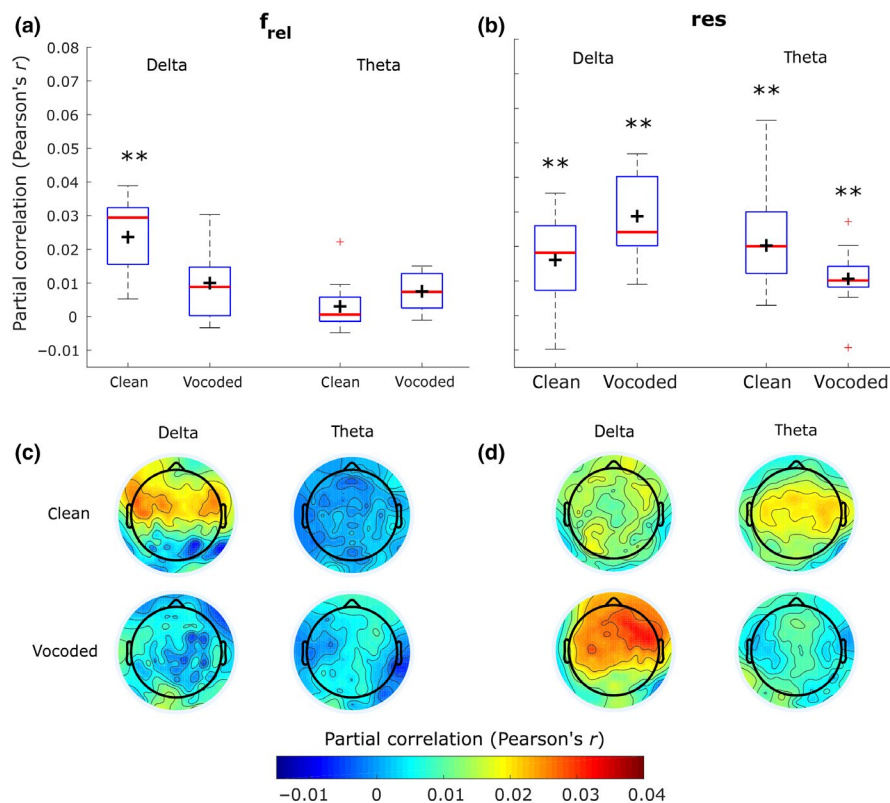


FIGURE 6 Partial correlations showing how well our pitch-related measures extracted from clean speech predict EEG that was recorded as subjects listened to clean and vocoded versions of speech snippets. Group-level distributions (average across best 30 channels; but using all channels produced the same pattern of results) are shown for delta and theta phase for (a) f_{rel} and (b) res . Black crosses indicate mean across all subjects, and asterisks indicate significance (one-tailed permutation test; $*p < .05$, $**p < .01$). The topographic distributions of these partial correlations are also shown in (c) and (d). [Colour figure can be viewed at wileyonlinelibrary.com]

speech segments. Partial correlation coefficients are shown in Figure 6. Consistent with our above results, we found significant tracking of f_{rel} in delta band for clean speech (one-tailed permutation test; $p = 9.99\text{e-}4$), but no tracking of this measure in vocoded speech ($p = .1918$). There was also no significant tracking of f_{rel} in either clean or vocoded speech in theta band ($p = .9411$ and $p = .1918$, respectively). In contrast, res was present in clean and vocoded speech in both frequency bands (delta: $p = 9.99\text{e-}4$ and $p = 9.99\text{e-}4$, theta: $p = 9.99\text{e-}4$ and $p = .0020$, respectively). In delta band, vocoded speech was tracked significantly better than clean speech (two-tailed Wilcoxon signed-rank, $z = 2.3412$, $p = .0192$); the reverse trend was observed in theta band although this was not significant ($z = -1.852$, $p = .0640$).

3.4 | No evidence for hemispheric lateralization of our pitch-related measures

Whilst there is consensus that speech processing occurs bilaterally, the differential roles of the two hemispheres remain controversial. Several auditory processing models (Poeppel, 2003; Zatorre, Belin, & Penhune, 2002) posit (or are compatible with the view) that the right hemisphere is relatively specialized for spectral information in contrast with the left hemisphere's specialization for rapidly fluctuating temporal cues. Certain aspects of low-level pitch processing have indeed been found to be lateralized to the right hemisphere (Griffiths, Johnsrude,

Dean, & Green, 1999; Johnsrude, Penhune, & Zatorre, 2000; Zatorre, Evans, & Meyer, 1994), but lesion and neuroimaging studies have suggested a less clear-cut picture with respect to prosody-related pitch (refer to Discussion). With that in mind, we tested for hemispheric lateralization in the processing of our pitch-related measures. To do so, we averaged the partial correlations of the delta-phase EEG predicted by our pitch-related features with the actual delta-phase EEG (i.e. corresponding to the results in Figure 4) over right and left hemispheres (39 on each side; midline and vertex electrodes were disregarded) for each subject and compared their distributions (Figure 7). No significant differences were observed between hemispheres for either pitch-related measure (two-tailed Wilcoxon signed-rank, f_{rel} : $z = 1.0865$, $p = .1365$; res : $z = 0.9658$, $p = .3341$).

4 | DISCUSSION

Recent fMRI and ECoG studies have shown that there are cortical pitch regions which respond primarily to resolved harmonics (Norman-Haignere et al., 2013) and that speaker-normalized relative pitch is encoded in high-gamma activity in cortex (Tang et al., 2017). Here, we looked to isolate indices corresponding to these pitch-processing phenomena in scalp-recorded low-frequency EEG during listening to continuous narrative speech. We computed and regressed two pitch-related features—harmonic resolvability (res) and relative pitch (f_{rel})—to ongoing

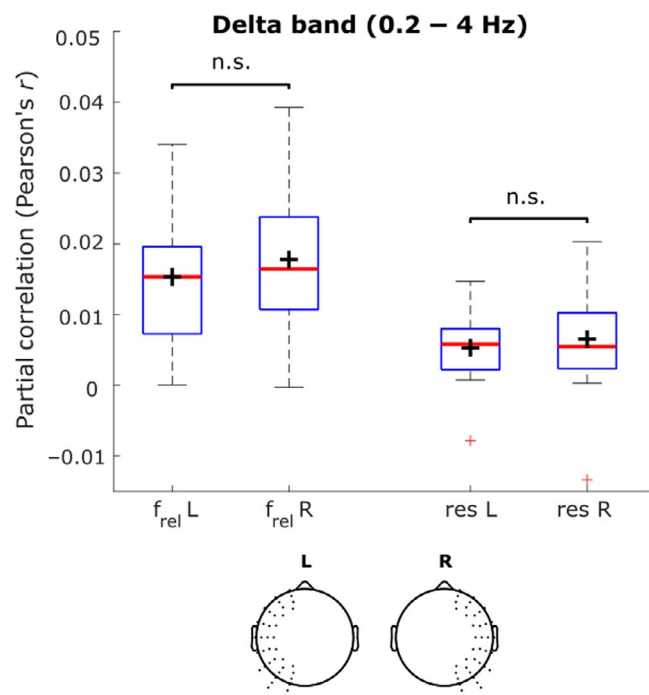


FIGURE 7 We tested for hemispheric dominance in the processing of pitch features by comparing the average of their partial correlations in delta band across left- and right-sided channels (midline and vertex channels were removed). No significant differences were observed between hemispheres for either of the features (two-tailed Wilcoxon signed-rank test (R vs. L); f_{rel} : $z = 1.4890$, $p = .1365$; res : $z = 0.9658$, $p = .3341$). [Colour figure can be viewed at wileyonlinelibrary.com]

EEG activity (broadband and analytic components of various frequency bands). To determine whether they contain unique predictive power, we accounted for three previously established acoustic and phonetic representations: the temporal envelope, a commonly used measure of the stimulus that captures the coherent fluctuations in time of multiple acoustic features, the spectrogram, which has been shown to model acoustic energy beyond envelope, and phonetic features, which captures some of the acoustics as well as higher-level categorical information.

We found evidence for the encoding of our *res* feature in broadband EEG and in its delta- and theta-phase analytic components. Combining this measure with our other acoustic/phonetic features to predict broadband EEG led to a significant improvement over using only the joint acoustic/phonetic model. Moreover, partial correlation analysis found that this feature contributed unique predictive power on a group level in broadband EEG, as well as in delta and theta phase. It was particularly prominent in theta phase, where it was significantly greater than chance for the majority of subjects. The question then arose as to whether the unique contribution of our *res* measure reflects pitch-specific activity. This was a concern particularly because we found the time course of the *res* measure to be highly correlated with

the temporal envelope. Indeed, it is evident from the large disparity in the prediction accuracies of *res* before and after controlling for other measures that the tracking of this measure overlaps markedly with these other features. And, given that the acoustic features we included are almost certainly imperfect representations of all coherent low-level fluctuations in the stimulus, it could be that the unique predictive power of *res* is simply due to it capturing some of these residual shared fluctuations. Furthermore, we found this measure to be tracked in both EEG recorded as subjects listened to clean speech and its noise-vocoded counterpart (which has degraded pitch). In fact, it was significantly more predictive of EEG for noise-vocoded speech than clean speech in delta phase. This adds to the complexity of pinpointing precisely what the tracking of this measure means.

We were initially motivated to test the *res* measure based on compelling fMRI findings (Norman-Haignere et al., 2013). Even so, our inability to clearly disambiguate this measure is in no way contradictory to their results. For one, they were able to isolate small, pitch-sensitive regions of auditory cortex on the basis of fMRI's excellent spatial resolution. EEG is lacking in that respect, but we were hoping to leverage EEG's superior temporal resolution to the same effect. Unfortunately, this proved to be complicated simply because many features of speech—*res* being one of them—covary in time. It may well be that the unique contribution of *res* to EEG prediction that we have shown contains some pitch-specific activity, but it appears to be too confounded to be a useful measure for indexing pitch processing per se. Nonetheless, it could still be potentially useful in applications where it is imperative to extract as much stimulus-related signal as possible from EEG.

With regard to our relative pitch measure, combining it with the joint acoustic/phonetic model led to a significant improvement in prediction accuracy in broadband EEG. It also portrayed a topographic distribution of prediction accuracies that was markedly different from that of the other measures. And, unlike acoustic/phonetic features and resolvability, the unique contribution of relative pitch was found to be band-limited to delta phase. The suprasegmental information carried by relative pitch in speech typically fluctuates over intervals of several hundred milliseconds to seconds, so delta-phase tracking is consistent with parsing and integrating information at this rate. To validate that this measure is reflective of pitch-specific activity, we compared the tracking of this measure in EEG when subjects listened to clean and noise-vocoded speech. As in experiment 1, we found significant tracking of relative pitch in delta band for clean speech; but, crucially, there was no tracking for vocoded speech. Together, these results give us confidence that this feature indexes pitch-specific processing.

Our finding complements that of Tang et al. (2017), showing that relative pitch is also reflected in low-frequency EEG phase. This is non-trivial as studies examining both phase-locked low-frequency (event-related potentials, ERPs, which

are somewhat analogous to the TRF) and non-phase-locked high-gamma EEG/ECOG activity (Crone, Boatman, Gordon, & Hao, 2001; Crone, Sinai, & Korzeniewska, 2006; Edwards et al., 2009; Engell & McCarthy, 2011) have found different spatial and temporal profiles for the two measures, suggesting that they result from distinct physiological mechanisms. Although the unique contribution of f_{rel} to predicting EEG appears to be small relative to overall acoustic/phonetic tracking, this was not unexpected as Tang et al. (2017) also found only a small percentage of electrodes within auditory cortex to be tuned to relative pitch in comparison with those tuned to the phonetic content of sentences.

A lot of EEG/MEG speech research in recent years has focused on the neural tracking of the temporal envelope and its functional roles (Ding & Simon, 2014). The temporal envelope of speech represents the slow amplitude fluctuations in the acoustic signal and conveys important markers for segmenting linguistic features (syllables in particular) (Rosen, 1992). Indeed, behavioural studies have shown that the envelope contains information essential for speech intelligibility (Drullman, Festen, & Plomp, 1994a, 1994b). Given the myriad cues represented in the temporal envelope, the precise function(s) of the observed envelope tracking response—whether it simply reflects passive general auditory encoding or specific language-related processing as well—remains debated. There has nonetheless been some empirical evidence that the measure correlates with speech comprehension (Ahissar et al., 2001; Molinaro & Lizarazu, 2018; Peelle, Gross, & Davis, 2013), suggesting that it relates to both acoustic–phonetic and linguistic processing. Our results here, showing a pitch-specific measure that is dissociable from the envelope, as well as the spectrogram and phonetic features, posit a parallel neural channel for processing meaning conveyed by prosodic pitch.

General auditory processing models (Poeppel, 2003; Zatorre et al., 2002) predict a right hemispheric bias for pitch processing. But neural correlates of prosodic pitch have only been observed to be right lateralized within the context of certain tasks, and typically when lexical content is no longer identifiable (Baum & Pell, 1999; Gandour et al., 2004; Kotz et al., 2003; Kreitewolf et al., 2014; Meyer, Alter, & Friederici, 2003; Meyer et al., 2002; Peretz, 1990; Plante et al., 2002; Sammler et al., 2015; Witteman et al., 2011; Zatorre, Evans, Meyer, & Gjedde, 1992). Left lateralization has also been observed within the context of contrasting tonal and non-tonal language processing (Gandour et al., 2004). Zatorre and Gandour (2008) surmised that multiple local asymmetries underlie prosodic pitch processing, with right hemispheric activity reflecting lower-level, domain-independent pitch processing (in accordance with general models) and left-hemispheric activity reflecting higher-level language-related processing. The emergence of lateralization effects is therefore dependent on the interaction of these

processes in a specific listening situation. Here, in a scenario where subjects listened to continuous natural (non-tonal) speech, much like they would do in real life, we did not find any hemispheric bias for relative pitch (nor for *res*, for that matter). That is, there was no significant difference in how accurately relative pitch could predict delta-phase EEG activity from the right and left hemispheres. Whilst it could be the case that the effect is simply too small to be detected with EEG given its low spatial resolution, our result lends support to the notion that prosodic pitch processing in the context of natural speech is subserved by a complex network spanning both hemispheres (Zatorre & Gandour, 2008).

Like ECoG, EEG has excellent millisecond temporal resolution that lends itself to the task of understanding how rapid speech dynamics are encoded, but with the added advantages of being portable and non-invasive. These properties facilitate the application of any findings to cognitively controlled smart hearing devices (Lunner, Rudner, & Rönnerberg, 2009) and speech processing studies in specific cohorts. For instance, individuals with autism often display disordered or unusual prosody (McCann & Peppé, 2003). Neural processing of prosody in individuals with autism remains an under-researched topic and an isolated index of relative pitch processing in continuous natural speech as shown here could help shed light on the nature of their prosodic deficits.

ACKNOWLEDGEMENTS

The authors would like to thank S. Norman-Haignere for sharing code for computing resolvability, and G. Di Liberto for the data. This study was supported by an Irish Research Council Government of Ireland Postgraduate Scholarship (GOIPG/2015/3378).

CONFLICT OF INTEREST

The authors declare no conflicts of interest.

DATA ACCESSIBILITY

Data for experiment 1 of this study are available on <https://datadryad.org/resource/doi:10.5061/dryad.070jc> (Natural Speech data set). Data for experiment 2 are available upon request.

AUTHOR CONTRIBUTIONS

The study was conceived by E.C.L. E.S.T. and M.S.C. analysed the data. E.S.T. wrote the manuscript. E.C.L. and E.S.T. edited the manuscript.

ORCID

Emily S. Teoh  <https://orcid.org/0000-0002-6351-1807>

REFERENCES

- Ahissar, E., Nagarajan, S., Ahissar, M., Protopapas, A., Mahncke, H., & Merzenich, M. M. (2001). Speech comprehension is correlated with temporal response patterns recorded from auditory cortex. *Proceedings of the National Academy of Sciences*, 98(23), 13367–13372. <https://doi.org/10.1073/pnas.201400998>
- Aiken, S. J., & Picton, T. W. (2008). Human Cortical Responses to the Speech Envelope. *Ear and Hearing*, 29(2), 139–157. <https://doi.org/10.1097/AUD.0b013e31816453dc>
- Baum, S. R., & Pell, M. D. (1999). The neural bases of prosody: Insights from lesion studies and neuroimaging. *Aphasiology*, 13(8), 581–608. <https://doi.org/10.1080/026870399401984>
- Boersma, P. (1993). Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound. In IFA Proceedings 17. (pp. 97–110).
- Boersma, P., & Weenink, D. 2016. Praat: doing Phonetics by Computer, Version 6.0.20. Retrieved from <http://www.fon.hum.uva.nl/praat/>
- Bolinger, D., & Bolinger, D. L. M. (1986). *Intonation and Its Parts: Melody in Spoken English*. Stanford, CA: Stanford University Press.
- Briley, P. M., Breakey, C., & Krumbholz, K. (2013). Evidence for pitch chroma mapping in human auditory cortex. *Cerebral Cortex (New York, NY)*, 23(11), 2601–2610. <https://doi.org/10.1093/cercor/bhs242>
- Broderick, M. P., Anderson, A. J., Liberto, G. M. D., Crosse, M. J., & Lalor, E. C. (2018). Electrophysiological correlates of semantic dissimilarity reflect the comprehension of natural narrative speech. *Current Biology*, 28(5), 803–809.e3. <https://doi.org/10.1016/j.cub.2018.01.080>
- Cole, J., & Shattuck-Hufnagel, S. (2016). New methods for prosodic transcription: Capturing variability as a source of information. *Laboratory Phonology: Journal of the Association for Laboratory Phonology*, 7(1), 8. <https://doi.org/10.5334/labphon.29>
- Crone, N. E., Boatman, D., Gordon, B., & Hao, L. (2001). Induced electrocorticographic gamma activity during auditory perception. *Clinical Neurophysiology*, 112(4), 565–582. [https://doi.org/10.1016/S1388-2457\(00\)00545-9](https://doi.org/10.1016/S1388-2457(00)00545-9)
- Crone, N. E., Sinai, A., & Korzeniewska, A. (2006). High-frequency gamma oscillations and human brain mapping with electrocorticography. In C. Neuper, & W. Klimesch (Eds.), *Progress in brain research*, Vol. 159 (pp. 275–295). London, UK: Elsevier. [https://doi.org/10.1016/S0079-6123\(06\)59019-3](https://doi.org/10.1016/S0079-6123(06)59019-3)
- Crosse, M. J., Di Liberto, G. M., Bednar, A., & Lalor, E. C. (2016). The Multivariate Temporal Response Function (mTRF) toolbox: A MATLAB toolbox for relating neural signals to continuous stimuli. *Frontiers in Human Neuroscience*, 10, <https://doi.org/10.3389/fnhum.2016.00604>
- Davis, M. H., & Johnsrude, I. S. (2003). Hierarchical Processing in Spoken Language Comprehension. *Journal of Neuroscience*, 23(8), 3423–3431. <https://doi.org/10.1523/JNEUROSCI.23-08-03423.2003>
- Di Liberto, G. M., Crosse, M. J., & Lalor, E. C. (2018). Cortical measures of phoneme-level speech encoding correlate with the perceived clarity of natural speech. *ENeuro*, 5(2). <https://doi.org/10.1523/eneuro.0084-18.2018>
- Di Liberto, G. M., O'Sullivan, J. A., & Lalor, E. C. (2015). Low-frequency cortical entrainment to speech reflects phoneme-level processing. *Current Biology*, 25(19), 2457–2465. <https://doi.org/10.1016/j.cub.2015.08.030>
- Ding, N., & Simon, J. Z. (2014). Cortical entrainment to continuous speech: Functional roles and interpretations. *Frontiers in Human Neuroscience*, 8, <https://doi.org/10.3389/fnhum.2014.00311>
- Drullman, R., Festen, J. M., & Plomp, R. (1994a). Effect of reducing slow temporal modulations on speech reception. *The Journal of the Acoustical Society of America*, 95(5 Pt 1), 2670–2680.
- Drullman, R., Festen, J. M., & Plomp, R. (1994b). Effect of temporal envelope smearing on speech reception. *The Journal of the Acoustical Society of America*, 95(2), 1053–1064. <https://doi.org/10.1121/1.408467>
- Edwards, E., Soltani, M., Kim, W., Dalal, S. S., Nagarajan, S. S., Berger, M. S., & Knight, R. T. (2009). Comparison of time-frequency responses and the event-related potential to auditory speech stimuli in human cortex. *Journal of Neurophysiology*, 102(1), 377–386. <https://doi.org/10.1152/jn.90954.2008>
- Engell, A. D., & McCarthy, G. (2011). The relationship of gamma oscillations and face-specific ERPs recorded subdurally from occipitotemporal cortex. *Cerebral Cortex (New York, NY)*, 21(5), 1213–1221. <https://doi.org/10.1093/cercor/bhq206>
- Gandour, J., Tong, Y., Wong, D., Talavage, T., Dziedzic, M., Xu, Y., ... Lowe, M. (2004). Hemispheric roles in the perception of speech prosody. *NeuroImage*, 23(1), 344–357. <https://doi.org/10.1016/j.neuroimage.2004.06.004>
- Giraud, A.-L., & Poeppel, D. (2012). Cortical oscillations and speech processing: Emerging computational principles and operations. *Nature Neuroscience*, 15(4), 511–517. <https://doi.org/10.1038/nn.3063>
- Gorman, K., Howell, J., & Wagner, M. (2011). Prosodylab-aligner: A tool for forced alignment of laboratory speech. *Canadian Acoustics*, 39(3), 192–193.
- Greenwood, D. D. (1961). Auditory masking and the critical band. *The Journal of the Acoustical Society of America*, 33(4), 484–502. <https://doi.org/10.1121/1.1908699>
- Griffiths, T. D., Johnsrude, I., Dean, J. L., & Green, G. G. (1999). A common neural substrate for the analysis of pitch and duration pattern in segmented sound? *NeuroReport*, 10(18), 3825–3830.
- Griffiths, T. D., Kumar, S., Sedley, W., Nourski, K. V., Kawasaki, H., Oya, H., ... Howard, M. A. (2010). Direct recordings of pitch responses from human auditory cortex. *Current Biology*, 20(12), 1128–1132. <https://doi.org/10.1016/j.cub.2010.04.044>
- Hall, D. A., & Plack, C. J. (2009). Pitch processing sites in the human auditory brain. *Cerebral Cortex*, 19(3), 576–585. <https://doi.org/10.1093/cercor/bhn108>
- Johnsrude, I. S., Penhune, V. B., & Zatorre, R. J. (2000). Functional specificity in the right human auditory cortex for perceiving pitch direction. *Brain*, 123(1), 155–163. <https://doi.org/10.1093/brain/123.1.155>
- Kotz, S. A., Meyer, M., Alter, K., Besson, M., von Cramon, D. Y., & Friederici, A. D. (2003). On the lateralization of emotional prosody: An event-related functional MR investigation. *Brain and Language*, 86(3), 366–376. [https://doi.org/10.1016/S0093-934X\(02\)00532-1](https://doi.org/10.1016/S0093-934X(02)00532-1)
- Kreitewolf, J., Friederici, A. D., & von Kriegstein, K. (2014). Hemispheric lateralization of linguistic prosody recognition in comparison to speech and speaker recognition. *NeuroImage*, 102, 332–344. <https://doi.org/10.1016/j.neuroimage.2014.07.038>
- Lalor, E. C., & Foxe, J. J. (2010). Neural responses to uninterrupted natural speech can be extracted with precise temporal resolution. *European Journal of Neuroscience*, 31(1), 189–193. <https://doi.org/10.1111/j.1460-9568.2009.07055.x>

- Lunner, T., Rudner, M., & Rönnerberg, J. (2009). Cognition and hearing aids. *Scandinavian Journal of Psychology*, 50(5), 395–403. <https://doi.org/10.1111/j.1467-9450.2009.00742.x>
- McCann, J., & Peppé, S. (2003). Prosody in autism spectrum disorders: A critical review. *International Journal of Language & Communication Disorders*, 38(4), 325–350. <https://doi.org/10.1080/1368282031000154204>
- McDermott, J. H., & Simoncelli, E. P. (2011). Sound texture perception via statistics of the auditory periphery: Evidence from sound synthesis. *Neuron*, 71(5), 926–940. <https://doi.org/10.1016/j.neuron.2011.06.032>
- Meyer, M., Alter, K., & Friederici, A. (2003). Functional MR imaging exposes differential brain responses to syntax and prosody during auditory sentence comprehension. *Journal of Neurolinguistics*, 16(4), 277–300. [https://doi.org/10.1016/S0911-6044\(03\)00026-5](https://doi.org/10.1016/S0911-6044(03)00026-5)
- Meyer, M., Alter, K., Friederici, A. D., Lohmann, G., & von Cramon, D. Y. (2002). fMRI reveals brain regions mediating slow prosodic modulations in spoken sentences. *Human Brain Mapping*, 17(2), 73–88. <https://doi.org/10.1002/hbm.10042>
- Molinaro, N., & Lizarazu, M. (2018). Delta (but not theta)-band cortical entrainment involves speech-specific processing. *European Journal of Neuroscience*, 48(7), 2642–2650. <https://doi.org/10.1111/ejn.13811>
- Norman-Haignere, S., Kanwisher, N., & McDermott, J. H. (2013). Cortical pitch regions in humans respond primarily to resolved harmonics and are located in specific tonotopic regions of anterior auditory cortex. *The Journal of Neuroscience*, 33(50), 19451–19469. <https://doi.org/10.1523/JNEUROSCI.2880-13.2013>
- Peelle, J. E., Gross, J., & Davis, M. H. (2013). Phase-locked responses to speech in human auditory cortex are enhanced during comprehension. *Cerebral Cortex*, 23(6), 1378–1387. <https://doi.org/10.1093/cercor/bhs118>
- Peretz, I. (1990). Processing of local and global musical information by unilateral brain-damaged patients. *Brain*, 113(Pt 4), 1185–1205.
- Plante, E., Creusere, M., & Sabin, C. (2002). Dissociating sentential prosody from sentence processing: Activation interacts with task demands. *NeuroImage*, 17(1), 401–410. <https://doi.org/10.1006/nimg.2002.1182>
- Poeppl, D. (2003). The analysis of speech in different temporal integration windows: Cerebral lateralization as ‘asymmetric sampling in time’. *Speech Communication*, 41(1), 245–255. [https://doi.org/10.1016/S0167-6393\(02\)00107-3](https://doi.org/10.1016/S0167-6393(02)00107-3)
- Rosen, S. (1992). Temporal information in speech: Acoustic, auditory and linguistic aspects. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 336(1278), 367–373. <https://doi.org/10.1098/rstb.1992.0070>
- Sammler, D., Grosbras, M.-H., Anwander, A., Bestelmeyer, P. E. G., & Belin, P. (2015). Dorsal and ventral pathways for prosody. *Current Biology*, 25(23), 3079–3085. <https://doi.org/10.1016/j.cub.2015.10.009>
- Shannon, R. V., Zeng, F. G., Kamath, V., Wygonski, J., & Ekelid, M. (1995). Speech recognition with primarily temporal cues. *Science*, 270(5234), 303–304.
- Steinhauer, K., Alter, K., & Friederici, A. D. (1999). Brain potentials indicate immediate use of prosodic cues in natural speech processing. *Nature Neuroscience*, 2(2), 191–196. <https://doi.org/10.1038/5757>
- Tang, C., Hamilton, L. S., & Chang, E. F. (2017). Intonational speech prosody encoding in the human auditory cortex. *Science*, 357(6353), 797–801. <https://doi.org/10.1126/science.aam8577>
- Witteman, J., van Ijzendoorn, M. H., van de Velde, D., van Heuven, V. J. J. P., & Schiller, N. O. (2011). The nature of hemispheric specialization for linguistic and emotional prosodic perception: A meta-analysis of the lesion literature. *Neuropsychologia*, 49(13), 3722–3738. <https://doi.org/10.1016/j.neuropsychologia.2011.09.028>
- Zatorre, R. J., Belin, P., & Penhune, V. B. (2002). Structure and function of auditory cortex: Music and speech. *Trends in Cognitive Sciences*, 6(1), 37–46. [https://doi.org/10.1016/S1364-6613\(00\)01816-7](https://doi.org/10.1016/S1364-6613(00)01816-7)
- Zatorre, R. J., Evans, A. C., & Meyer, E. (1994). Neural mechanisms underlying melodic perception and memory for pitch. *The Journal of Neuroscience*, 14(4), 1908–1919.
- Zatorre, R. J., Evans, A. C., Meyer, E., & Gjedde, A. (1992). Lateralization of phonetic and pitch discrimination in speech processing. *Science*, 256(5058), 846–849.
- Zatorre, R. J., & Gandour, J. T. (2008). Neural specializations for speech and pitch: Moving beyond the dichotomies. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 363(1493), 1087–1104. <https://doi.org/10.1098/rstb.2007.2161>

How to cite this article: Teoh ES, Cappelloni MS, Lalor EC. Prosodic pitch processing is represented in delta-band EEG and is dissociable from the cortical tracking of other acoustic and phonetic features. *Eur J Neurosci*. 2019;50:3831–3842. <https://doi.org/10.1111/ejn.14510>