



Published in final edited form as:

J Neural Eng. ; 16(5): 056004. doi:10.1088/1741-2552/ab1e01.

Biometric identification of listener identity from frequency following responses to speech

Fernando Llanos¹, Zilong Xie², Bharath Chandrasekaran¹

¹Department of Communication Sciences and Disorders, University of Pittsburgh, Pittsburgh, PA 15213, USA

²Department of Hearing and Speech Sciences, University of Maryland, College Park, MD 20742, USA

Abstract

Objective.—We investigated the biometric specificity of the frequency following response (FFR), an EEG marker of early auditory processing that reflects phase-locked activity from neural ensembles in the auditory cortex and subcortex (Bidelman, 2015a, 2018; Chandrasekaran & Kraus, 2010; Coffey et al., 2017). Our objective is two-fold: demonstrate that the FFR contains information beyond stimulus properties and broad group-level markers, and to assess the practical viability of the FFR as a biometric across different sounds, auditory experiences, and recording days.

Approach.—We trained the hidden Markov model (HMM) to decode listener identity from FFR spectro-temporal patterns across multiple frequency bands. Our dataset included FFRs from twenty native speakers of English or Mandarin Chinese (10 per group) listening to Mandarin Chinese tones across three EEG sessions separated by days. We decoded subject identity within the same auditory context (same tone and session) and across different stimuli and recording sessions.

Main results.—The HMM decoded listeners for averaging sizes as small as one single FFR. However, model performance improved for larger averaging sizes (e.g., 25 FFRs), similarity in auditory context (same tone and day), and lack of familiarity with the sounds (i.e., native English relative to native Chinese listeners). Our results also revealed important biometric contributions from frequency bands in the cortical and subcortical EEG.

Significance.—Our study provides the first deep and systematic biometric characterization of the FFR and provides the basis for biometric identification systems incorporating this neural signal.

Keywords

biometric identification system; machine learning; frequency following response; electroencephalogram; hidden Markov model

1 Introduction

The scalp-recorded frequency-following response (FFR) is considered a potent electrophysiological marker of the integrity of *stimulus encoding* within the auditory system

(Bidelman, 2015a; Chandrasekaran & Kraus, 2010; Krishnan et al., 2005; Skoe & Kraus, 2010). The FFR reflects stimulus properties to such an extent that when the response is sonified, the content of the speech is highly intelligible (e.g., Galbraith et al., 1995). While the FFR is highly faithful to the stimulus properties, the extent of faithfulness is shaped by short-term and long-term (positive or negative) listening experiences. For example, prior work has demonstrated enhanced stimulus encoding fidelity (more faithful FFRs) in some groups relative to others (e.g., musicians relative to non-musicians; Bidelman et al., 2011a–b). The FFR is also shown to be less faithful in some groups (e.g. older adults relative to young adults, children with developmental dyslexia relative to neurotypical children; e.g., Skoe et al., 2013). Thus, the signal reconstructed by the auditory system reflects the properties of the evoking stimulus (typically measured as a correlation between the signal and the FFR), as well as group-level differences. In the current study we investigated an intriguing possibility: that in addition to reflecting the properties of the stimulus and group-level differences, the FFR also reflects the individual. From a practical perspective, we will also assess the robustness of the FFR as a biomarker across a broad range of conditions. In the next few sections we will review prior efforts using the EEG as a biometric identification system, discuss the FFR in more detail, and finally highlight the specifics of the current study.

1.1 Biometric identification systems

The EEG reflects electrical oscillations that relate to sensory and cognitive functioning (Luck, 2004) as well as oscillations that are specific to subject biometrics (Vogel, 2000). Previous research has taken advantage of the specificity of electrical oscillation patterns to discriminate subjects from EEGs (Del Pozo-Banos et al., 2014). The interest in EEG metrics for personal identity is as old as the earliest EEG experiments. The first EEG biometric studies aimed to identify heritability markers in family members, including monozygotic and dizygotic twins (e.g., Davis & Davis, 1936; Young, Lader, & Fenton, 1972; for a review see Vogel 1970, 2000). In these studies, genetically close subjects (e.g., monozygotic twins) exhibited greater similarity in the alpha and beta waves of the resting EEG than subjects that were genetically less similar (e.g., non-relatives). This finding has been further investigated in EEG modalities other than the resting state and rhythms other than alpha and beta (Van Beijsterveldt et al., 1996; Van Beijsterveldt & Van Baal, 2002; Vogel, 2000; Zietsch et al., 2007).

Since the first published EEG biometric identification system (Stassen, 1980), different studies have used cortical oscillation patterns to discriminate non-relatives across multiple conditions (Del Pozo-Banos et al., 2014). These conditions include the identification of subjects from EEGs elicited in the resting state with the eyes closed (Poulos et al., 2002) and open (Paranjape et al., 2001), motor tasks (Yang & Deravi, 2012), brain imaginary tasks (Bao, Wang & Hu, 2009; Xiao & Hu, 2010), and complex cognitive tasks involving mathematical operations and the discrimination of visual patterns (Palaniappan, 2005, 2006; Ravi & Palaniappan, 2005). The results of this bulk of research are not always consistent with respect to the frequency band and scalp montage providing the best biometric markers. In general, the optimal combination of parameters varies as a function of the task and features to decode (Del Pozo-Banos et al., 2014; Eischen, Luckritz & Polich, 1995; van

Beijsterveldt & van Baal, 2002). Interestingly, a recent biometric identification study informed by multiple EEG tasks has reported a task-independent subject-discrimination plateau starting at approximately 30 Hz, in the gamma range (Del Pozo-Banos et al., 2015).

The models and features used to discriminate subjects from cortical EEGs are also diverse, from simple linear classifiers input with event-related potentials (ERPs) selected ad hoc, such as the alpha peak or the P300 wave (Gupta, Palaiappan & Paramesran, 2012; van Beijsterveldt & van Baal, 2002), to complex data-driven classifiers trained with vectors of power spectral density (PSD) or autoregressive coefficients (Mohammadi et al., 2006; Nguyen, Tran, Huang, & Sharma, 2012; Poulos, Rangoussi, & Alexandris, 1999; Singhal & RamKumar, 2007). Here, data-driven classifiers tend to discriminate subjects better than hypothesis-driven features. This result suggests that discriminant information is sparse across multiple EEG dimensions.

1.2 The frequency following response

A common denominator of the previous EEG biometric identification studies is that -to the best of our knowledge- they were exclusively informed by cortical responses. Here, we examined the biometric specificity of the FFR, a scalp-recorded electrophysiological potential that reflects sustained phase-locking activity primarily from cortical and subcortical neural ensembles (Coffey et al., 2016; Coffey, Musacchia, & Zatorre, 2017; Jewett & Williston, 1971; Moushegian, Rupert, & Stillman, 1973; Sohmer, Pratt, & Kinarti, 1977). The FFR reflects changes in scalp electrical potentials due to rapid neural synchronization to stimulus periodicity over time, thus providing a temporal coding of complex acoustic properties that underlies auditory processing (Chandrasekaran & Kraus, 2010; Chandrasekaran, Skoe, & Kraus, 2014) (see Fig. 1).

While the primary source-generators of the FFR are thought to be in the rostral brainstem (Bidelman, 2018; Chandrasekaran & Kraus, 2010; Marsh, Brown & Smith, 1974), more recent evidence is suggestive of cortical sources as well (Coffey et al., 2016, 2017). Neural activity in subcortical structures can also be modulated by cortical structures via descending (or efferent) projections (Kral & Eggermont, 2017; Keuroghlian & Knudsen, 2007). Efferent projections play a role in enhancing the subcortical encoding of the acoustic properties that are behaviorally-relevant to the organism (Andéol et al., 2011; Malmierca, Anderson & Antunes, 2014). Therefore, the FFR also reflects functional aspects of auditory processing that go beyond the neural transduction of stimulus properties (e.g., selective attention, Xie, Reetzke & Chandrasekaran, 2018; and predictive coding, Luo, Wong, & Chandrasekaran, 2016; Skoe & Chandrasekaran, 2014).

The effects of long-term auditory experience in the FFR are well documented in native speakers of tonal languages, like Mandarin Chinese (Bidelman, Gandour, & Krishnan, 2011a, 2011b; Krishnan et al., 2005; Krishnan et al., 2004; Weiss & Bidelman, 2015). Mandarin Chinese speakers are systematically exposed to pitch modulations that are functionally relevant in their soundscape. For example, in Mandarin Chinese, the syllable /ma/ can be interpreted as ‘mother’, ‘horse’, ‘hemp’, or ‘scold’ depending on whether is pronounced with a high-level, low-dipping, low-rising, or high-falling fundamental frequency (F0) (Gandour, 1994) (see Fig. 2). Critically, FFRs from native

Chinese speakers reflect a more faithful neural encoding of Mandarin tones than FFRs from native English speakers (Bidelman, Gandour, & Krishnan, 2011a, 2011b; Krishnan et al., 2009; Krishnan et al., 2005).

In practice, the FFR is extracted from raw EEG responses with a latency of approximately 10 ms and a band-pass filter, from 80 to 1000 Hz, that roughly reflects the phase-locking range of neurons in the rostral brainstem (Chandrasekaran & Kraus, 2010; Skoe & Kraus, 2010; Krishnan, 2002; Moushegian et al., 1973). Within this frequency range, the FFR captures stimulus' spectral patterns with remarkable fidelity (see Fig. 1) (Bidelman & Krishnan, 2009; Greenberg et al., 1987). Previous research has taken advantage of this property of the FFR to decode sound categories from FFR responses (Llanos, Xie, & Chandrasekaran, 2017; Reetzke et al., 2018; Yi et al., 2017). Building upon this finding, we aimed to decode listeners from FFRs as a proxy to assess the biometric specificity of the FFR.

Practically, the FFR presents several features that could be attractive for an EEG biometric system. First, the FFR can be easily recorded with a simple montage of three electrodes (Aiken & Picton, 2008; Skoe & Kraus, 2010). While cortical responses can also be recorded with a few scalp electrodes (e.g., Krishnan et al., 2012), the discrimination of subjects from cortical EEGs typically requires more than three electrodes (e.g., Abdullah et al., 2010). Second, relative to cortical responses, the FFR is less affected by changes in listeners' cognitive states (e.g., Varghese, Bharadwaj, & Shinn-Cunningham, 2015). As a result, it is typically recorded while listeners are resting or watching a silent movie. Such procedure simplifies the data acquisition process and reduces the number of myogenic artifacts (Akhoun et al., 2008; Skoe & Kraus, 2010). Relatedly, in contrast to cortical responses, the FFR is quite robust against ocular artifacts, as these artifacts oscillate at frequencies below the typical FFR bandwidth. Third, the FFR is highly stable across EEG sessions separated by days (Bidelman et al., 2018; Xie et al., 2017), months (Reetzke, et al., 2018; Song, Nicol & Kraus, 2011), or even years (Hornickel, Knowles & Kraus, 2012). This property may facilitate the identification of subjects over time. Fourth, the FFR can be recorded easily in neonates, and across the lifespan. Finally, the FFR reflects a wide range of inter-subject variability due to subtle individual differences in auditory processing (Chandrasekaran, Kraus, & Wong, 2012; Coffey et al., 2017; Coffey et al., 2016; Hornickel et al., 2009. Ruggles, Bharadwaj, & Shinn-Cunningham, 2012) and auditory experience (Chandrasekaran, Krishnan, & Gandour, 2007; Krishnan, Gandour, & Bidelman, 2010; Krishnan et al., 2009; Wong et al., 2007). This property may also facilitate the discrimination of subjects from FFRs.

1.3 The present study

In the current study, we examined the biometric specificity of the FFR. To this end, we used a machine learning classifier to decode subject identity from a previously published FFR dataset from our lab (Xie et al., 2017). In this dataset, FFRs were collected in three (repeated) sessions separated by days from native speakers of Mandarin-Chinese and native speakers of English. In each session, participants listened to multiple repetitions of two Mandarin-Chinese tones. The use of behaviorally-relevant stimuli (e.g., Palaniappan &

Raveendran, 2002), or behaviorally-relevant tasks (Marcel & Millan, 2007), is not uncommon for the biometric identification practice. Behaviorally-relevant materials are expected to engage neural processing across multiple cognitive and neural domains and thus provide a wide variety of biometric markers. In previous FFR studies, the use of linguistically-relevant stimuli served to evoke functional markers of cognitive processing that were beyond the mere sensory encoding of acoustic properties (e.g., language-dependent markers; Krishnan et al., 2005).

We decoded subject identity from native and non-native speakers of Mandarin Chinese to investigate the role of long-term auditory experience on the FFR biometrics. We investigated the extent to which the biometric specificity of the FFR is permeable to long-term auditory experience. This question was promoted by research findings showing that the stimulus fidelity of the FFR is enhanced by long-term auditory experience (e.g., Krishnan et al., 2005; Xie et al., 2017). Specifically, we assessed whether listeners familiar with the evoking stimulus (native Chinese-listeners) were better decoded than listeners unfamiliar with the evoking stimulus (native English-listeners). To further assess the effects of long-term auditory experience on the FFR biometrics, we also examined whether native speakers of the same language (either English or Chinese) were biometrically more similar than native speakers of different languages.

We used the performance of a hidden Markov model (HMM), trained to decode subjects from FFRs, as a metric to quantify the biometric specificity of the FFR. The HMM captures short-term dependencies in complex time series (Juang, & Rabiner, 1991; Rabiner, 1990) and is therefore suitable to model complex time-varying signals, like the FFR. In a previous study (Llanos et al., 2017), we used the HMM to decode Mandarin tone categories from FFRs recorded from native speakers of Mandarin Chinese and English. Our results yielded a significantly greater tone decoding accuracy for the Mandarin Chinese group. This finding suggests that the HMM can pick up on subtle, biologically-relevant differences that could be useful to discriminate subjects.

We decoded subject identity in the same auditory context (i.e., same tone and EEG session) and across different auditory contexts (i.e., different EEG sessions and different Mandarin tones). Since the goal of a biometric identification system is to authenticate personal identity over time, we decoded the same subject identity across different EEG sessions to assess the temporal stability of biometric profiles. We also decoded subject identity across different Mandarin tones to investigate the presence of pitch-free biometric markers. Finally, we addressed two further questions relevant for biometric identification purposes. To increase the signal-to-noise ratio (SNR) of the FFR, FFRs are averaged across multiple stimulus repetitions (Kraus & Skoe, 2010). Since large averaging sizes are more time consuming than small sizes because they require a greater number of experimental trials, our first question was on the smallest averaging size required to identify subjects. The second question was on the EEG bands providing the best discriminant information. Since the cortical components of the FFR manifest at frequencies lower than approximately 200 Hz (Bidelman, 2018), and the FFR bandwidth goes up to 1 kHz, this analysis helped us to weight the biometric contributions of cortical and subcortical oscillation patterns. With our results, we expect to provide the first deep and systematic biometric characterization of the FFR.

2 Methods

2.1 Dataset

Our FFR dataset (Xie et al., 2017) included EEGs from ten adult native speakers of Mandarin Chinese and ten adult native speakers of American English with no prior language experience with Chinese or any other tonal language. We included no participants with formal instruction in music to minimize the effect of music experience on the FFR (Bidelman et al., 2011a, 2011b). Participants completed three independent EEG sessions with no task interpolated between them. The average number of days between the first two EEG sessions was 1.25 days (ranging from 1 to 3 days). The third EEG session was always administered the day after the second session. In each session, participants listened to 1000 repetitions of the same Chinese syllable (/a/, 256 ms) pronounced with a high-level Mandarin tone and a low-rising Mandarin tone. Stimulus F₀-contours ranged from 100 to 130 Hz (see Fig. 2B). Auditory stimuli were blocked by tone category and presented with alternate polarities to attenuate the cochlear microphonic effect (Skoe & Kraus, 2010). Participants were instructed to ignore the sounds and focus on a silent movie of choice.

EEGs were acquired with same system (Brain Vision ActiCHamp), software (Pycorder 1.0.7), and scalp montage across sessions. The scalp montage consisted of three Ag-AgCl scalp electrodes connected to a 50-gain pre-amplifier from the left mastoid (ground), right mastoid (reference), and Cz (active). We used the vertex electrode to enhance the representation of neural activity from the rostral brainstem and attenuate the cochlear microphonic effect (Skoe & Kraus, 2010). The use of a mastoid reference is consistent with our previous work (Xie et al., 2017) and prior work from other labs (Bidelman & Krishnan, 2010; Krishnan et al. 2010). This vertical montage provided us with good SNR levels for FFR averaging sizes smaller or equal to those documented in the FFR literature (e.g., 1000 trials).

Single-trial FFRs were pre-processed from raw EEGs with customized scripts written in MATLAB (R2017a). We band-pass filtered the EEG channel from 80 to 1000 Hz using a second-order zero-phase Butterworth filter. Single-trial FFRs were segmented from the filtered channel using a latency of 10 ms and the duration of the evoking stimulus (256 ms). Single-trial FFRs were baseline corrected by subtracting the mean EEG magnitude of the baseline noise (from -40 to 0 ms) and baseline corrected trials with a maximum absolute magnitude higher than 35 μ V were rejected as potential artifacts. Since previous studies (e.g., Del Pozo-Banos et al., 2015) have identified biometric contributions from EEG bandwidths below the typical FFR bandwidth (80 – 1000 Hz), we kept a low-pass filtered (<80 Hz) version of each unrejected trial for analysis purposes.

2.2 Decoding of subject identity

Each subject identity (N=20) was decoded with a different HMM, customized in MATLAB (R2017a). Each HMM was trained with Viterbi (Levinson, Rabiner, & Sondhi, 1983; Lou, 1995) to recognize FFRs from a single subject. The HMM was structured as a chain of three hidden states feedforward connected to the following states. This number of states was chosen to account for biometric patterns at the onset, mid and offset portions of the FFR. To

increase the SNR of the FFR, training and testing sets were sub-averaged with a moving average window including a balanced number of responses to different stimulus polarity items (see Fig. 3). We computed the short-term fast-Fourier transform (ST-FFT) of each sub-averaged FFR in decibels. Then, we kept the first 13 coefficients of the discrete-cosine transform (DCT) of each ST-FFT time frame. This number of coefficients allowed us to reduce input dimensionality while preserving the overall spectral shape of each ST-FFT frame (Davis & Mermelstein, 1990; Nadeu, Macho & Hernando, 2001). Each sequence of short-term DCT frames was encoded into a sequence of HMM emissions using the same codebook. The codebook was created by clustering all the DCT frames in all training sets. We used a k-means approach (Lloyd, 1982) to split the distribution of DCT frames into a series of Voronoi cells, one cell for each centroid in the k-means cluster. Each sequence of DCT frames was then quantized into a series of discrete HMM emissions by mapping each DCT frame in the sequence into the identification number of its corresponding cell in the DCT space. Each HMM was trained with the FFRs included in the training set of the corresponding subject and tested with the FFRs conveyed by the testing sets of all subjects.

The overlap between consecutive ST-FFT frames was set to 75%. This overlap has been shown to capture changes in FFR spectral-quality over time without consuming too much computational time (Llanos et al., 2017; Krishnan et al., 2005; Krishnan et al., 2009). The other model parameters were estimated with an optimization method. These parameters were the length of the ST-FFT frame (10, 20, or 40 ms), the codebook length (50, 100, or 150 codewords), the training size (500, 750, or 875 FFRs, out of 1000), and the FFR averaging size (50, 100, or 200 trials). We applied a first-order iterative method to find the combination of parameters that maximized model performance for the high-level tone and the first EEG session; we used the FFRs to the other tone and sessions to evaluate the stability of biometric profiles across different auditory contexts (e.g., over time and across sounds). We initialized the optimization method with a frame length of 40 ms, a codebook length of 50 codewords, a training size of 500 FFRs, and an averaging size of 200 trials. These parameters were adapted from a previous study using the HMM to decode Mandarin Chinese tones from FFRs (Llanos et al., 2017). We updated these parameters in subsequent iterations by searching for the frame length, codebook length, training size, and averaging size that maximized subject decoding performance when the other three parameters were fixed. We continued updating the parameters until they converged to a local maximum.

Model performance was scored as the area under the receiver operating characteristic curve (AUROC). The receiver operating characteristic (ROC) curve models the trading between true positives (y-axis) and false alarms (x-axis) across different detection thresholds. Improvements in identification rate are characterized by larger true positive rates over false alarm rates across threshold levels, leading to larger AUROCs (Bradley, 1997). The detection threshold was defined as the smallest likelihood accepted to classify an FFR into the subject class of the HMM. We created a ROC curve for each subject-specific model by varying the detection threshold across multiple levels spanning the likelihood range of the HMMs. Then, we plotted the true positive rate (i.e., the proportion of testing trials above threshold) against the false alarm rate (i.e., the proportion of testing trials below threshold) for each level. The ROC curve is currently a standard metric of model performance in machine learning and data mining work (Streiner & Cairney, 2007; Swets, Dawes &

Monahan, 2000), including machine learning studies with a focus on clinical diagnosis (Linden, 2006). While no metric is perfect, ROC analysis presents some properties of interest, relative to classic accuracy analysis. First, accuracy analysis only controls for type I errors, whereas ROC analysis simultaneously controls for type I and II errors. As a result, ROC analysis helps us to estimate model performance across different potential scenarios, where the investigator may choose boosting the true positive rate at the expense of the false alarm rate or vice versa. In addition, the level of chance in accuracy analysis decreases as the sample size increases, thus increasing the probability of the same model being over chance. In contrast, the level of chance in ROC analysis (AUROC = 0.5) remains stable across sample sizes. Since the level of chance is a common standard in machine learning approaches to cognitive neuroscience (e.g., Bohland et al., 2012; Feng et al., 2017), ROC analysis provides a fairer chance level comparison across studies using different sample sizes.

One limitation of the ROC analysis is that true positives and false alarms are computed for each model independently. While this approach is very fruitful to calibrate the detection threshold that minimizes the trade-off cost between true positives and false alarms for each subject-specific model (single-subject recognition designs), it does not offer a straightforward interpretation of model performance in a multi-subject classification design, where the same subject is classified across different models. To provide a more comprehensive picture of model performance in this context, we also examined the accuracy of a multi-subject classifier informed by the performance of each subject-specific HMM. Here, each FFR in the testing sets was classified into subject class of the HMM providing the highest likelihood. We used the accuracy of the classifier as a metric of model performance in a multi-subject classification design.

2.3 Analyses

We trained the HMM to decode a different subject at a time (20 subjects total). To assess the subject specificity of the FFR within the same auditory context, the HMM was trained and tested with FFRs to the high-level tone in the first session (i.e., these were the tone and session used to optimized model parameters). We also used these FFRs to investigate the performance of the model in a multi-subject classification design, the effects of long-term auditory experience, averaging size, and EEG bandwidth on subject identification. We used the FFRs to the other tone and in the other sessions to investigate the stability of biometric profiles over time and across different sounds. In this case, the HMM was trained and tested across different sessions and tones (i.e., across different auditory contexts).

The optimal model parameters obtained with the optimization method were 20 ms (ST-FFT frame length), 150 codewords (codebook length), 750 FFRs (training size, out of 1000 FFRs), and 200 trials (averaging size). These parameters were not changed across analyses, except when we investigated the effects of averaging size on model performance. In this case, averaging size was systematically manipulated to answer the corresponding research question. In a similar vein, the HMM was input with spectral coefficients within the standard FFR bandwidth (80 Hz - 1 kHz), except when we examined the amount of biometric information conveyed at specific frequency bands. In this case, the HMM was also input

with spectral coefficient below 80 Hz, in the cortical EEG range. The number of cross-validation folds was also fixed across analyses. Since the training size was 250 FFRs (= 1000 FFRs - training size), the performance of the HMM was cross-validated with four folds.

2.3.1 Subject specificity and auditory experience—We used the following rating scale (Carter, Pan & Galandjuk, 2016) to categorize the performance of the HMM: *excellent* (AUROC > 0.9), *good* (> 0.8), *fair* (> 0.7), *poor* (> 0.6), or *bad* (> 0.5). To assess the effects of long-term auditory experience with the evoking stimuli on subject identification, we compared Chinese and English AUROCs with a two-sample t-test. To assess whether native speakers of the same language were biometrically more similar than native speakers of different languages, we examined the distribution of language profiles (English and Chinese) in the same biometric space. To create this space, we classified all testing FFRs into the subject class of the HMMs that maximized their likelihood function. Then, we arranged English and Chinese biometric profiles into the same two-dimensional scaling space as a function of the relative distance of their models in the confusion matrix provided by the classification process. We clustered the biometric profiles in the biometric space using an unsupervised clustering of Gaussian mixtures (full and shared covariance; McLachlan, 2000). The number of mixtures, from 1 to 5, was selected with the Davies-Bouldin criterion (Davies & Bouldin, 1979). Then, we used the Shannon entropy formula to quantify the level of language dispersion in each cluster and on average (the lower the Entropy, the lower the dispersion; Shannon, 1948). Finally, we estimated the probability of sampling a lower mean dispersion from a chance-level distribution of 1000 mean dispersion values obtained by the random permutation of language labels across subjects. The resulting probability served us as a p-value to reject the hypothesis that the mean dispersion value obtained in first place was due to chance.

2.3.2 Cross-stimulus and temporal stability—To assess the stability of biometric profiles across different sounds, we trained each HMM (one per subject) with FFRs to the high-level tone. Then, we evaluated the model with novel FFRs to the high-level tone and the low-rising tone. We compared the AUROCs in each evaluation set (i.e., the high-level tone vs. the low-rising tone) with a linear mixed-effects model with set and subject as fixed and random effects, and the high-level tone as the reference level. To assess the temporal stability of biometric profiles, we trained each HMM (one per subject) with FFRs to the high-level tone in the first EEG session. Then, we evaluated the model with FFRs to the same tone in the first, second, and third sessions. To compare the AUROCs in the second and third EEG sessions with the AUROCs in the first session, we used a linear-mixed effects model with session and subject as fixed and random effects, and the first session as the reference level.

2.3.3 Averaging size and EEG bandwidth—To investigate the effects of FFR averaging size on subject decoding, we input each HMM (one per subject) with training and testing sets sub-averaged across different sizes (200, 50, 25, 10, 5, or 1 trial). We focused on small averaging sizes because small sizes require a lower number of stimulus repetitions and, therefore, they can be used to identify subjects faster, in the order of few seconds. We

assessed the effects of the averaging size on model performance with a linear-mixed effects model with size and subject as fixed and random effects, and the largest size (200 trials) as the reference level. To estimate the amount of biometric information encoded in each EEG bandwidth, the HMMs (one per subject) were input with DCT coefficients from a different frequency band at a time. Frequency bands below the FFR bandwidth were segmented as follows: delta (1 – 4 Hz), theta (4 – 8 Hz), alpha (8 – 12 Hz), beta (12 – 30 Hz), low gamma (30 – 60 Hz) and high gamma (60 – 100). Frequency bands higher than 100 Hz, in the FFR bandwidth, were segmented in 10 steps of 100 Hz ranging from 100 Hz to 1 kHz (100 – 200 Hz, 200 – 300 Hz, etc.). We used steps of 100 Hz because they captured the main modulations of the DCT spectral envelope of the FFR (see the DCT spectral envelopes shown in Fig. 3). To assess the effects of frequency band on model performance, we used a linear-mixed effects model with band and subject as fixed and random effects, and the band providing the largest AUROC as the reference level (60 – 100 Hz).

3 Results

3.1 Subject specificity and auditory experience

Within the same auditory context (i.e., same tone and session), the HMM decoded subject identity with a mean AUROC of 0.93, in the excellent range (> 0.9 ; dotted line in Fig. 4A). The two-sample t-test (see bar plot in Fig. 4A) revealed a statistically significant difference between language groups ($t[10] = -2.53$, $p = 0.03$). Here, native speakers of English ($M = 0.97$, $SD = 0.02$) were decoded with higher AUROCs than native speakers of Chinese (Chinese: $M = 0.89$, $SD = 0.09$). This result indicates that model performance was modulated by language experience. Specifically, listeners that were more familiar with the target stimuli (i.e., Chinese listeners) were harder to decode.

Since Mandarin-Chinese tones are more faithfully represented in Chinese FFRs, relative to English FFRs (Xie et al., 2017), it is possible that the Chinese targets were more likely to be confused with one another by the model because their FFRs exhibited greater convergence to stimulus F0 than the English targets. To examine this possibility, we used multidimensional scaling analysis to project Chinese and English FFRs into the same space as a function of their relative Euclidean distances. Since FFR properties often require thousands of repetitions to consolidate in the averaged response, we maximized the averaging size to 1000 trials, rather than using the testing size (250 trials). Then, we calculated the level of dispersion around each language group mean (SD) in the multidimensional scaling space. Our results revealed that the Chinese FFRs were more concentrated around their language group mean ($SD = 2.1$) in the multidimensional scaling space than the English FFRs ($SD=3$). This result brings support to the hypothesis that Chinese FFRs were more similar to one another than English FFRs.

The analysis of language dispersion in the HMM biometric space (see Fig. 4B) revealed a quite small level of language dispersion across biometric clusters (mean Shannon entropy = 0.18, from 0 to 1). This result indicates that most of the subjects within the same cluster happened to be native speakers of the same language. Furthermore, the results of the permutation analysis indicated that this level of language dispersion was not likely due to chance ($p < 0.01$), as more than the 99% of the mean dispersion levels sampled from the

chance-level distribution were higher than 0.18. In the HMM biometric space, the Chinese group was also more concentrated around its language group mean ($SD = 2.8$) than the English group ($SD = 3.2$). In addition, while seven out of ten English subjects were spread around the left margin of the HMM biometric space, away from the scope of a considerably number of subjects, seven out of ten Chinese subjects were concentrated in the middle area of the biometric space, flanked by a higher density of proximal biometric profiles. Hypothetically, these Chinese subjects in the middle of the biometric space may have penalized the performance of the model with a lower true-positive rate. Indeed, their mean AUROC ($M = 0.83$) was smaller than the mean AUROC of the Chinese subjects located in the right margin of the biometric space ($M = 0.98$). Altogether, these factors may have contributed to impair recognition in the Chinese group.

In addition to the analysis of ROC curves, we also examined the performance of the HMM in a multi-subject classification design. Here, the HMM classifier classified subjects with a mean accuracy of 0.74 ($SD = 0.22$), way above the level of chance ($= 0.05$). This level of classification accuracy is consistent with prior EEG biometric work (Del-Pozo-Banos et al., 2015; Singhal & RamKumar, 2007; Yeom, Suk, & Lee, 2013), although it is still smaller than the 0.9 level of accuracy achieved in other EEG biometric studies (e.g., Palaniappan & Raveendran, 2002). Mean classification accuracy was higher for the English models ($M = 0.8$), relative to the Chinese models ($M = 0.67$), consistently with the results of the ROC analyses. Altogether, these results indicate that the FFR contains information that is relevant for the classification of multiple subjects.

3.2 Cross-stimulus and temporal stability

Model performance decreased when the HMM was trained and tested across different Mandarin tones (see Fig. 4C). Here, model performance declined from excellent AUROCs (>0.9) to good AUROCs (>0.8) when the HMM was trained and tested with FFRs to different tones ($M = 0.83$; $SD = 0.16$) rather than with FFRs to the same tone ($M = 0.93$, $SD = 0.08$) ($t [38] = -3.07$, $p = 0.003$). Model performance also declined when the HMM was trained and tested across different EEG sessions (see Fig. 4D). Here, model performance decreased from excellent to fair AUROCs when the HMM was trained and tested with FFRs from different EEG session (second session: $M = 0.76$, $SD = 0.24$; third session: $M = 0.77$, $SD = 0.24$) rather than the same session (first session: $M = 0.93$, $SD = 0.08$; $-2.95 < t [57] < -2.86$, $ps < 0.005$).

3.3 Averaging size and frequency band

Model performance decreased with averaging size (see Fig. 5A), but here the decline was quite small, going from excellent to less excellent or good AUROCs. In particular, the AUROC decreased from 0.93 (200 trials, $SD = 0.08$) to 0.92 (50 trials, $SD = 0.06$), 0.91 (25 trials, $SD = 0.06$), 0.89 (10 trials, $SD = 0.07$), 0.88 (5 trials, $SD = 0.08$), and 0.85 (1 trial, $SD = 0.07$). The linear mixed-effects model revealed that only averaging sizes lower than 25 trials (10, 5 and 1 trial) were statistically different than the reference level (200 trials; $-6.8 < t_{Stat}[114] < -4.16$, $ps < 0.002$). The analysis of model performance by frequency band (see Fig. 5B–C) revealed two frequency bands with AUROCs as large as the reference level (reference-level band: 60 – 100 Hz). These bandwidths were: 400 – 500 Hz and 100 – 200

Hz. The other bands provided AUROCs that were statistically smaller than the reference level ($-4.2 < t_{\text{Stat}[285]} < -2.3$, $p < 0.01$), although some of these AUROCs were still close to the reference level (e.g., the AUROCs of 800 – 900 Hz and 500 – 600 Hz).

4 Discussion

We trained the HMM to decode subject identity from FFRs to Mandarin tones recorded from native speakers of Mandarin Chinese and native speakers of English naïve to lexical tones. We used the performance of the HMM as a metric to characterize the biometric specificity of the FFR for the same auditory context (i.e., same Mandarin tone and EEG session) and across different auditory contexts (i.e., across different EEG sessions or Mandarin tones). Since the FFR is modulated by long-term auditory experiences, we also investigated the effects of stimulus familiarity on subject identification. Here, we aimed to determine if native Chinese listeners are better decoded than native English listeners, and if native speakers of the same language are biometrically more similar than native speakers of different languages. We also aimed to identify the smallest FFR averaging needed to decode subject identity and the EEG frequency bands providing the best biometric information.

4.1 Subject specificity

In the same auditory context, listeners were decoded with excellent and good AUROCs and classified with accuracy scores way above chance. This result indicates that, in addition to the high fidelity to stimulus characteristics, the FFR also contains patterns that are subject specific. How these subject-specific patterns relate to subtle individual differences in auditory processing is a question that goes beyond the scope of the present study, as we only manipulated the degree of auditory experience with the evoking stimuli. Our results suggest that at least part of these patterns might not be directly related to the neural encoding of stimulus properties (e.g., F0 contours). This claim is supported by the following findings. First, while the FFR typically requires a large averaging size to reflect stimulus-related properties (typically more than 1000 trials; Skoe & Kraus, 2010), our models performed well for averaging sizes as small as one single trial. Second, our results revealed important contributions to subject decoding performance from frequency bands that were above and below the stimulus F0 range. Finally, while the FFR reflects a more robust neural encoding of pitch in native Chinese-speakers, relative to native English-speakers (Krishnan et al., 2005), our results yielded better identification scores for the group of native English speakers. These findings could be a consequence of the type of features that we used to train and evaluate the HMM. Instead of using pitch-relevant properties, such as F0 contours, we input the HMM with linguistically-blind spectral coefficients. A visual inspection of our DCT envelopes in the time domain (see Fig. 6) shows that our models were more informed by signal properties related to amplitude modulation (e.g., the analytic signal) than frequency modulation (e.g., F0 contours).

4.2 Auditory experience

Listeners unfamiliar with the evoking stimuli (i.e., native English speakers) were *better decoded* than listeners familiar with the evoking stimuli (i.e., native Chinese speakers). This result shows that subject identification can be affected by sound familiarity. Our results also

suggest that subject-dependent neural plasticity is permeable to language experience. While native speakers of the same native language (either English or Chinese) were grouped into more than one cluster in the biometric space, they were not mixed with the native speakers of the other language (with the exception of one listener, out of 20). The fact that native speakers of different languages were almost never mixed in the same cluster suggests that the distribution of biometric profiles was modulated by language experience at some extent. Otherwise, our results should have revealed a greater level of language mixture across clusters; instead, the results of the permutation analysis showed that the amount of language dispersion across clusters was statistically smaller than would be expected if language did not play a role.

In the study where we collected the FFR dataset used here (Xie et al., 2017), Mandarin-Chinese listeners showed a more faithful neural encoding of Mandarin tones than native speakers of English. Specifically, the F0 contours of the FFRs in the Mandarin-Chinese group were more similar to the F0 contour of the evoking stimulus than in the English group. This result was consistent with prior FFR work showing a language-dependent enhancement of neural pitch encoding (e.g., Krishnan et al., 2005). While our results indicate that subject-related plasticity is not uniquely driven by the neural encoding of pitch, it is possible that Chinese listeners were more difficult to decode because their FFRs were more similar. However, the fact that Chinese and English clusters were interspersed in the biometric space suggests that the distribution of biometric profiles was also driven by variables unrelated to language experience; otherwise, our results should have shown two clearly differentiated clusters (one for each language group).

4.3 Cross-stimulus and temporal stability

The HMM decoded listener identity across different sounds and days. This finding indicates that the FFR contains biometric features that are stable over time and across sounds. Since we decoded listener identity across Mandarin tones, they demonstrate that the FFR contains biometric information that is not directly related to the encoding of stimulus F0. However, the performance of the HMM across days and tones was not as good as for the same auditory context. This result is consistent with the literature for cortical EEGs (e.g., Del Pozo-Banos et al., 2015), and suggests that the FFR must also contain biometric patterns that are not stable across stimuli and over time.

4.4 Averaging size and frequency band

The impact of the FFR averaging size on the HMM was quite small. Here, the HMM decoded subject identity with good scores for averaging sizes as small as one single trial. This result indicates that the biometric specificity of the FFR is quite robust against poor SNR levels, as the SNR of the FFR decreases with averaging size (Skoe & Kraus, 2010). Subject identification was also modulated by the EEG bandwidth. Here, our results revealed contributions from frequency bands above and below the phase-locking limitations of neurons in the auditory cortex (approximately at 200 Hz). This finding demonstrates the presence of EEG biometric patterns in the subcortical range. We also found biometric contributions from frequency bands in the (high) gamma range (<100 Hz). This finding is

consistent with the results of a previous EEG biometric identification study reporting a subject discrimination plateau starting at mid gamma (Del Pozo-Banos et al., 2015).

4.5 Biometric identification systems

Our results suggest that EEG biometric identification systems could benefit from including the FFR in their metrics. As we noted above, the FFR can be collected with a simple EEG montage of three electrodes. Furthermore, our results for the averaging size suggest that subjects could be identified very quickly with very few stimulus repetitions. A quick simulation of this hypothetical situation using only the first 10 testing trials of each HMM (single-trial averaging size) yielded a mean AUROC ($M = 0.86$) that was quite similar to the mean AUROC obtained with a testing set of 1000 FFRs ($M = 0.85$). Another property that could be attractive for an EEG biometric identification system is that FFRs can be simultaneously collected with cortical ERPs (e.g., auditory evoked potentials: Bidelman et al. 2015b). Thus, FFRs and cortical ERPs could be integrated into the same system to increase the number of markers used to authenticate subjects.

If the goal is to improve the authentication of subjects across different days, we recommend training a full-bandwidth model (1 – 1000 Hz) with FFRs from different days. A quick simulation of this setting revealed an improvement from 0.77 (trained with the first session, tested with the third session) to 0.82 (trained with the first two sessions, tested with the third session). Based on our results, we also recommend training and testing the model with the same evoking stimuli. Since stimulus familiarization has a negative impact on model performance, we suggest using unfamiliar stimuli to boost decoding accuracy. Here, the evoking stimuli could be replaced after a while to avoid the negative effects of ongoing stimulus familiarization after several identification sessions.

With our results, we expect to set a preliminary basis for future studies using data driven metrics to retrieve complex FFR markers of individual differences in neurotypical and clinical populations. The FFR contains markers of concussion occurrence and severity in mild traumatic injury populations; e.g., children with a concussion elicit smaller and slower neural responses than neurotypical children (Kraus et al., 2016). Since machine learning models can pick up on a wider and more complex selection of signal properties than hypothesis-driven models (e.g., models in which the features of interest are selected ad hoc), data-driven approaches to EEG biometrics could contribute to refine the subject specificity of the current EEG models of traumatic brain injury. One research question that we could not address in our study, but that may be worthwhile to investigate, is how individual differences in subject decoding accuracy trade off with individual differences in auditory processing; e.g., do individuals with hearing impairment exhibit a more differential EEG biometric profile? Another question that remains open is how our results would scale with increasing sample size. Our sample size ($N = 20$) could be considered a standard sample size in the FFR literature. This sample size is similar, or larger, than the ones reported in a substantial number of EEG biometric identification studies (e.g., Sadanao, Miyamoto, & Nakanishi, 2008, Palaniappan, 2005; Poulos, Rangoussi, & Kafetzopoulos, 1998), but smaller than the 48 subjects reported in Campisi et al., (2011) or the 82 subjects reported in Stassen (1989). While the FFR has received increased attention in the last decade, the

number of FFR datasets currently available is still significantly smaller than the number of cortical EEG datasets. We hope that, as the field progresses, we will count with larger datasets to test our systems.

References

- Abdullah MK, Subari KS, Loong JLC, & Ahmad NN (2010). Analysis of the EEG signal for a practical biometric system. *World Academy of Science, Engineering and Technology*, 68, 1123–1127.
- Aiken SJ, & Picton TW (2008). Envelope and spectral frequency-following responses to vowel sounds. *Hearing research*, 245(1–2), 35–47. [PubMed: 18765275]
- Akhoun I, Gallégo S, Moulin A, Ménard M, Veuillet E, Berger-Vachon C, & Thai-Van H (2008). The temporal relationship between speech auditory brainstem responses and the acoustic pattern of the phoneme /ba/ in normal-hearing adults. *Clinical Neurophysiology*, 119(4), 922–933. [PubMed: 18291717]
- Bao X, Wang J, & Hu J (2009). Method of individual identification based on electroencephalogram analysis. In *International conference on new trends in information and service science, NISS'09* (pp. 390–393).
- Begleiter H, & Porjesz B (2006). Genetics of human brain oscillations. *International Journal of Psychophysiology*, 60(2), 162–171. [PubMed: 16540194]
- Bidelman GM (2015a). Multichannel recordings of the human brainstem frequency-following response: scalp topography, source generators, and distinctions from the transient ABR. *Hearing research*, 323, 68–80. [PubMed: 25660195]
- Bidelman GM (2015b). Towards an optimal paradigm for simultaneously recording cortical and brainstem auditory evoked potentials. *Journal of neuroscience methods*, 241, 94–100. [PubMed: 25561397]
- Bidelman GM (2018). Subcortical sources dominate the neuroelectric auditory frequency-following response to speech. *Neuroimage*, 175, 56–69. [PubMed: 29604459]
- Bidelman GM, & Krishnan A (2009). Neural correlates of consonance, dissonance, and the hierarchy of musical pitch in the human brainstem. *Journal of Neuroscience*, 29(42), 13165–13171. [PubMed: 19846704]
- Bidelman GM, Gandour JT, & Krishnan A (2011a). Cross-domain effects of music and language experience on the representation of pitch in the human auditory brainstem. *Journal of cognitive neuroscience*, 23(2), 425–434. [PubMed: 19925180]
- Bidelman GM, Gandour JT, & Krishnan A (2011b). Musicians and tone-language speakers share enhanced brainstem encoding but not perceptual benefits for musical pitch. *Brain and cognition*, 77(1), 1–10. [PubMed: 21835531]
- Bohland JW, Saperstein S, Pereira F, Rapin J, & Grady L (2012). Network, anatomical, and non-imaging measures for the prediction of ADHD diagnosis in individual subjects. *Frontiers in systems neuroscience*, 6, 78. [PubMed: 23267318]
- Bradley AP (1997). The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern recognition*, 30(7), 1145–1159.
- Campisi P, Scarano G, Babiloni F, DeVico Fallani F, Colonnese S, Maiorana E, et al. (2011). Brain waves based user recognition using the eyes closed resting conditions protocol. In *IEEE international workshop on information forensics an security, WIFS'11* (pp. 1–6).
- Carter JV, Pan J, Rai SN, & Galandiuk S (2016). ROC-ing along: Evaluation and interpretation of receiver operating characteristic curves. *Surgery*, 159(6), 1638–1645. [PubMed: 26962006]
- Chandrasekaran B, & Kraus N (2010). The scalp-recorded brainstem response to speech: Neural origins and plasticity. *Psychophysiology*, 47(2), 236–246. [PubMed: 19824950]
- Chandrasekaran B, Kraus N, & Wong PC (2012). Human inferior colliculus activity relates to individual differences in spoken language learning. *Journal of Neurophysiology*, 107(5), 1325–1336. [PubMed: 22131377]

- Chandrasekaran B, Krishnan A, & Gandour JT (2007). Mismatch negativity to pitch contours is influenced by language experience. *Brain research*, 1128, 148–156. [PubMed: 17125749]
- Chandrasekaran B, Skoe E, & Kraus N (2014). An integrative model of subcortical auditory plasticity. *Brain topography*, 27(4), 539–552. [PubMed: 24150692]
- Coffey EB, Chepesiuk AM, Herholz SC, Baillet S, & Zatorre RJ (2017). Neural correlates of early sound encoding and their relationship to speech-in-noise perception. *Frontiers in neuroscience*, 11, 479. [PubMed: 28890684]
- Coffey EB, Colagrosso EM, Lehmann A, Schönwiesner M, & Zatorre RJ (2016). Individual differences in the frequency-following response: relation to pitch perception. *PLoS one*, 11(3), e0152374. [PubMed: 27015271]
- Coffey EB, Herholz SC, Chepesiuk AM, Baillet S, & Zatorre RJ (2016). Cortical contributions to the auditory frequency-following response revealed by MEG. *Nature communications*, 7, 11070.
- Coffey EB, Musacchia G, & Zatorre RJ (2017). Cortical correlates of the auditory frequency-following and onset responses: EEG and fMRI evidence. *Journal of Neuroscience*, 37(4), 830–838. [PubMed: 28123019]
- Davies DL, & Bouldin DW (1979). A cluster separation measure. *IEEE transactions on pattern analysis and machine intelligence*, (2), 224–227. [PubMed: 21868852]
- Davis H, & Davis PA (1936). Action potentials of the brain: In normal persons and in normal states of cerebral activity. *Archives of Neurology & Psychiatry*, 36(6), 1214–1224.
- Davis SB, & Mermelstein P (1990). Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences. In *Readings in speech recognition* (pp. 65–74).
- Del Pozo-Banos M, Alonso JB, Ticay-Rivas JR, & Travieso CM (2014). Electroencephalogram subject identification: A review. *Expert Systems with Applications*, 41(15), 6537–6554.
- Del Pozo-Banos M, Travieso CM, Weidemann CT, & Alonso JB (2015). EEG biometric identification: a thorough exploration of the time-frequency domain. *Journal of neural engineering*, 12(5), 056019. [PubMed: 26394698]
- Eischen S, Luckritz J, & Polich J (1995). Spectral analysis of EEG from families. *Biological Psychology*, 41(1), 61–68. [PubMed: 8562674]
- Feng G, Gan Z, Wang S, Wong PC, & Chandrasekaran B (2017). Task-general and acoustic-invariant neural representation of speech categories in the human brain. *Cerebral cortex*, 28(9), 3241–3254.
- Galbraith GC, Arbagey PW, Branski R, Comerci N, & Rector PM (1995). Intelligible speech encoded in the human brain stem frequency-following response. *Neuroreport*, 6(17), 2363–2367. [PubMed: 8747154]
- Gallinat J, Winterer G, Herrmann CS, & Senkowski D (2004). Reduced oscillatory gamma-band responses in unmedicated schizophrenic patients indicate impaired frontal network processing. *Clinical Neurophysiology*, 115(8), 1863–1874. [PubMed: 15261865]
- Gandour JT (1994). Phonetics of tone. *The encyclopedia of language & linguistics*, 6, 3116–3123.
- Greenberg S, Marsh JT, Brown WS, & Smith JC (1987). Neural temporal coding of low pitch. I. Human frequency-following responses to complex tones. *Hearing research*, 25(2–3), 91–114. [PubMed: 3558136]
- Gupta C, Palaniappan R, & Paramesran R (2012). Exploiting the p300 paradigm for cognitive biometrics. *International Journal of Cognitive Biometrics*, 1(1), 6–38.
- Hornickel J, Skoe E, Nicol T, Zecker S, & Kraus N (2009). Subcortical differentiation of stop consonants relates to reading and speech-in-noise perception. *Proceedings of the National Academy of Sciences*, 106(31), 13022–13027.
- Jewett DL, & Williston JS (1971). Auditory-evoked far fields averaged from the scalp of humans. *Brain*, 94(4), 681–696. [PubMed: 5132966]
- Juang BH, & Rabiner LR (1991). Hidden Markov models for speech recognition. *Technometrics*, 33(3), 251–272.
- Krishnan A (2002). Human frequency-following responses: representation of steady-state synthetic vowels. *Hearing research*, 166(1), 192–201. [PubMed: 12062771]

- Krishnan A, Bidelman GM, Smalt CJ, Ananthakrishnan S, & Gandour JT (2012). Relationship between brainstem, cortical and behavioral measures relevant to pitch salience in humans. *Neuropsychologia*, 50(12), 2849–2859. [PubMed: 22940428]
- Krishnan A, Gandour JT, & Bidelman GM (2010). The effects of tone language experience on pitch processing in the brainstem. *Journal of Neurolinguistics*, 23(1), 81–95. [PubMed: 20161561]
- Krishnan A, Gandour JT, Bidelman GM, & Swaminathan J (2009). Experience dependent neural representation of dynamic pitch in the brainstem. *Neuroreport*, 20(4), 408. [PubMed: 19223793]
- Krishnan A, Xu Y, Gandour JT, & Cariani PA (2004). Human frequency-following response: representation of pitch contours in Chinese tones. *Hearing research*, 189(1), 1–12. [PubMed: 14987747]
- Krishnan A, Xu Y, Gandour JT, & Cariani PA (2005). Encoding of pitch in the human brainstem is sensitive to language experience. *Cognitive Brain Research*, 25(1), 161–168. [PubMed: 15935624]
- Levinson SE, Rabiner LR, & Sondhi MM (1983). An introduction to the application of the theory of probabilistic functions of a Markov process to automatic speech recognition. *Bell System Technical Journal*, 62(4), 1035–1074.
- Linden A (2006). Measuring diagnostic and predictive accuracy in disease management: an introduction to receiver operating characteristic (ROC) analysis. *Journal of evaluation in clinical practice*, 12(2), 132–139. [PubMed: 16579821]
- Llanos F, Xie Z, & Chandrasekaran B (2017). Hidden Markov modeling of frequency-following responses to Mandarin lexical tones. *Journal of neuroscience methods*, 291, 101–112. [PubMed: 28807860]
- Lloyd S (1982). Least squares quantization in PCM. *IEEE transactions on information theory*, 28(2), 129–137.
- Lou HL (1995). Implementing the Viterbi algorithm. *IEEE Signal processing magazine*, 12(5), 42–52.
- McLachlan G, & Peel D (2000). *Finite Mixture Models*. Hoboken, NJ: John Wiley & Sons, Inc.
- Mohammadi G, Shoushtari P, Molaee Ardekani B, & Shamsollahi MB (2006). Person identification by using AR model for EEG signals. Paper presented at the Proceeding of World Academy of Science, Engineering and Technology.
- Moushegian G, Rupert AL, & Stillman RD (1973). Scalp-recorded early responses in man to frequencies in the speech range. *Electroencephalography and clinical neurophysiology*, 35(6), 665–667. [PubMed: 4128165]
- Nadeu C, Macho D, & Hernando J (2001). Time and frequency filtering of filter-bank energies for robust HMM speech recognition. *Speech Communication*, 34(1–2), 93–114.
- Nguyen P, Tran D, Huang X, & Sharma D (2012). A proposed feature extraction method for eeg-based person identification. Paper presented at the Proceedings on the International Conference on Artificial Intelligence (ICAI).
- Palaniappan R (2005). Identifying individuality using mental task based brain-computer interface. In 3rd International conference on intelligent sensing and information processing, ICISIP'05, December 2005 (pp. 238–242).
- Palaniappan R (2006). Electroencephalogram signals from imagined activities: A novel biometric identifier for a small population. In Proceedings of the 7th international conference on intelligent data engineering and automated learning, IDEAL'06, September 2006 (pp. 604–611).
- Paranjape R, Mahovsky J, Benedicenti L, & Koles Z (2001). The electroencephalogram as a biometric. In Proceedings on Canadian conference on electrical and computer engineering, May 2001 (Vol. 2, pp. 1363–1366).
- Porjesz B, Rangaswamy M, Kamarajan C, Jones KA, Padmanabhapillai A, & Begleiter H (2005). The utility of neurophysiological markers in the study of alcoholism. *Clinical Neurophysiology*, 116(5), 993–1018. [PubMed: 15826840]
- Poulos M, Rangoussi M, & Alexandris N (1999). Neural network based person identification using EEG features. Paper presented at the Acoustics, Speech, and Signal Processing, 1999. Proceedings., 1999 IEEE International Conference on.
- Poulos M, Rangoussi M, & Kafetzopoulos E (1998). Person identification via the EEG using computational geometry algorithms In 9th European Signal Processing Conference (EUSIPCO 1998) (pp. 1–4). IEEE.

- Poulos M, Rangoussi M, Alexandris N, & Evangelou A (2002). Person identification from the EEG using nonlinear signal classification. *Methods of information in Medicine*, 41(1), 64–75. [PubMed: 11933767]
- Poulos M, Rangoussi M, Chrissikopoulos V, & Evangelou A (1999). Parametric person identification from the EEG using computational geometry. Paper presented at the Electronics, Circuits and Systems, 1999. Proceedings of ICECS'99. The 6th IEEE International Conference on.
- Rabiner LR (1990). A tutorial on hidden Markov models and selected applications in speech recognition. In *Readings in speech recognition* (pp. 267–296).
- Ravi K, & Palaniappan R (2005a). Recognizing individuals using their brain patterns. In 3rd International conference on information technology and applications, ICITA 2005, July 2005 (Vol. 2, pp. 520–523).
- Reetzke R, Xie Z, Llanos F, & Chandrasekaran B (2018). Tracing the Trajectory of Sensory Plasticity across Different Stages of Speech Learning in Adulthood. *Current Biology*, 28(9), 1419–1427. e1414. [PubMed: 29681473]
- Ruggles D, Bharadwaj H, & Shinn-Cunningham BG (2012). Why middle-aged listeners have trouble hearing in everyday settings. *Current Biology*, 22(15), 1417–1422. [PubMed: 22727697]
- Sadanao B, Miyamoto C, & Nakanishi I (2008, 7). Personal authentication using new feature vector of brain wave. In *ITC-CSCC: International Technical Conference on Circuits Systems, Computers and Communications* (pp. 673–676).
- Singhal GK, & RamKumar P (2007). Person identification using evoked potentials and peak matching. Paper presented at the Biometrics Symposium, 2007.
- Skoe E, & Chandrasekaran B (2014). The layering of auditory experiences in driving experience-dependent subcortical plasticity. *Hearing research*, 311, 36–48. [PubMed: 24445149]
- Skoe E, & Kraus N (2010). Auditory brainstem response to complex sounds: a tutorial. *Ear and hearing*, 31(3), 302. [PubMed: 20084007]
- Skoe E, Chandrasekaran B, Spitzer ER, Wong PC, & Kraus N (2014). Human brainstem plasticity: the interaction of stimulus probability and auditory learning. *Neurobiology of learning and memory*, 109, 82–93. [PubMed: 24291573]
- Skoe E, Krizman J, Anderson S, & Kraus N (2013). Stability and plasticity of auditory brainstem function across the lifespan. *Cerebral Cortex*, 25(6), 1415–1426. [PubMed: 24366906]
- Smith JC, Marsh JT, Greenberg S, & Brown WS (1978). Human auditory frequency-following responses to a missing fundamental. *Science*, 201(4356), 639–641. [PubMed: 675250]
- Sohmer H, Pratt H, & Kinarti R (1977). Sources of frequency following responses (FFR) in man. *Electroencephalography and clinical neurophysiology*, 42(5), 656–664. [PubMed: 67025]
- Stassen H (1980). Computerized recognition of persons by EEG spectral patterns. *Electroencephalography and Clinical Neurophysiology*, 49(12), 190–194. [PubMed: 6159164]
- Streiner DL, & Cairney J (2007). What's under the ROC? An introduction to receiver operating characteristics curves. *The Canadian Journal of Psychiatry*, 52(2), 121–128. [PubMed: 17375868]
- Swets JA, Dawes RM, & Monahan J (2000). Better decisions through science. *Scientific American*, 283(4), 82–87.
- Van Beijsterveldt C, & Van Baal G (2002). Twin and family studies of the human electroencephalogram: a review and a meta-analysis. *Biological psychology*, 61(1–2), 111–138. [PubMed: 12385672]
- Van Beijsterveldt C, Molenaar P, De Geus E, & Boomsma D (1996). Heritability of human brain functioning as assessed by electroencephalography. *American journal of human genetics*, 58(3), 562. [PubMed: 8644716]
- Varghese L, Bharadwaj HM, & Shinn-Cunningham BG (2015). Evidence against attentional state modulating scalp-recorded auditory brainstem steady-state responses. *Brain Research*, 1626, 146–164. [PubMed: 26187756]
- Vogel F (1970). The genetic basis of the normal human electroencephalogram (EEG). *Humangenetik*, 10(2), 91–114. [PubMed: 5528299]
- Vogel F (2000). *Genetics and the Electroencephalogram* (Vol. 16). Springer.

- Weiss MW, & Bidelman GM (2015). Listening to the brainstem: musicianship enhances intelligibility of subcortical representations for speech. *Journal of Neuroscience*, 35(4), 1687–1691. [PubMed: 25632143]
- Wong PC, Skoe E, Russo NM, Dees T, & Kraus N (2007). Musical experience shapes human brainstem encoding of linguistic pitch patterns. *Nature neuroscience*, 10(4), 420–422. [PubMed: 17351633]
- Xie Z, Reetzke RD, & Chandrasekaran B (2017). Stability and plasticity in neural encoding of linguistically-relevant pitch patterns. *Journal of Neurophysiology*, jn. 00445.02016.
- Yang S, & Deravi F (2012). On the effectiveness of EEG signals as a source of biometric information. In 3rd International conference on emerging security technologies, EST'12, September 2012 (pp. 49–52).
- Yazdani A, Roodaki A, Rezatofghi S, Misaghian K, & Setarehdan SK (2008). Fisher linear discriminant based person identification using visual evoked potentials. Paper presented at the Signal Processing, 2008. ICSP 2008. 9th International Conference on.
- Yeom SK, Suk HI, & Lee SW (2013). Person authentication from neural activity of face-specific visual self-representation. *Pattern Recognition*, 46(4), 1159–1169.
- Yi H-G, Xie Z, Reetzke R, Dimakis AG, & Chandrasekaran B (2017). Vowel decoding from single-trial speech-evoked electrophysiological responses: A feature-based machine learning approach. *Brain and Behavior*.
- Young JP, Lader MH, & Fenton GW (1972). A twin study of the genetic influences on the electroencephalogram. *Journal of Medical Genetics*, 9(1), 13. [PubMed: 5063515]
- Zhang X, Begleiter H, Porjesz B, & Litke A (1997). Electrophysiological evidence of memory impairment in alcoholic patients. *Biological Psychiatry*, 42(12), 1157–1171 [PubMed: 9426886]
- Zietsch BP, Hansen JL, Hansell NK, Geffen GM, Martin NG, & Wright MJ (2007). Common and specific genetic influences on EEG power bands delta, theta, alpha, and beta. *Biological psychology*, 75(2), 154–164. [PubMed: 17316957]

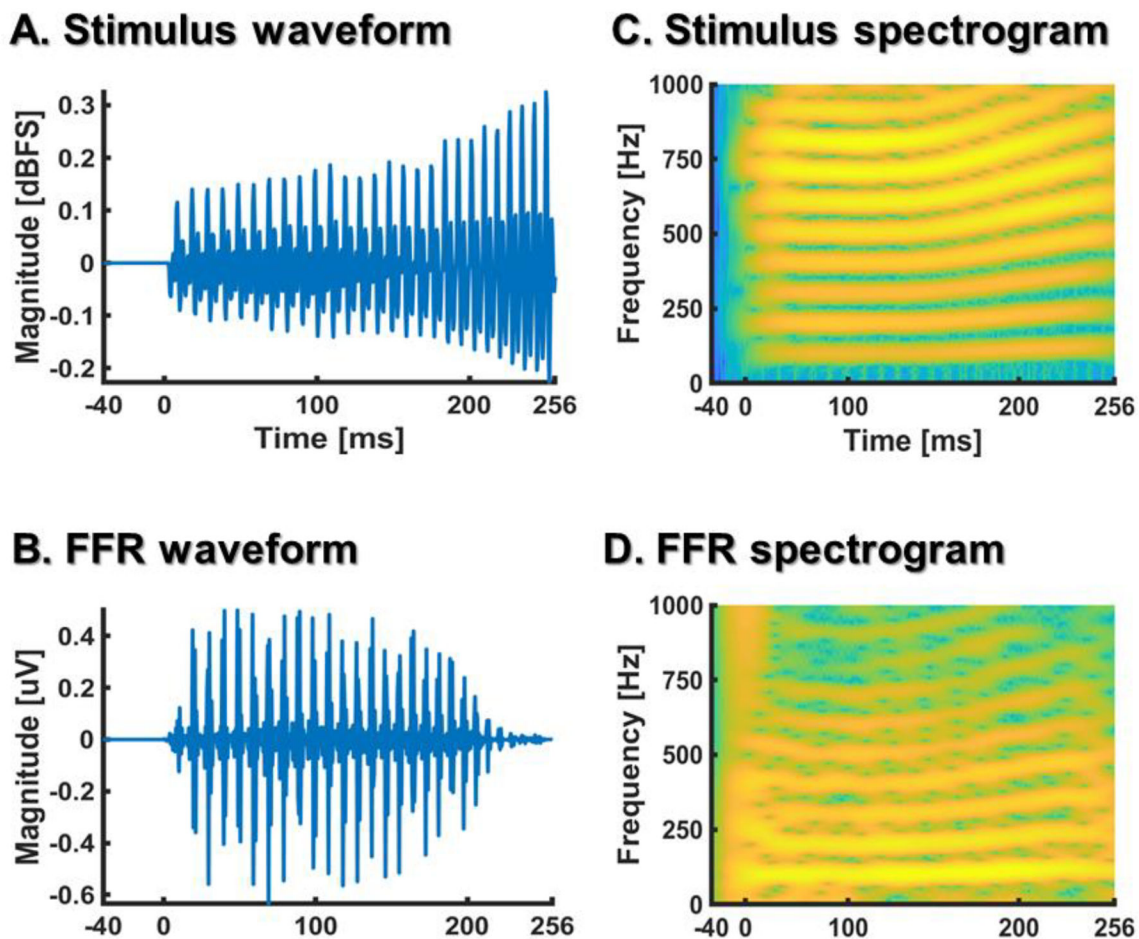


Figure 1. Waveforms (A, B) and spectrograms (C, D) of a rising Mandarin tone (A, C) and the evoked FFR (B, D) from a participant included in our study. The FFR was averaged across 1000 stimulus repetitions to leverage the signal-to-noise ratio.

MANDARIN CHINESE TONES

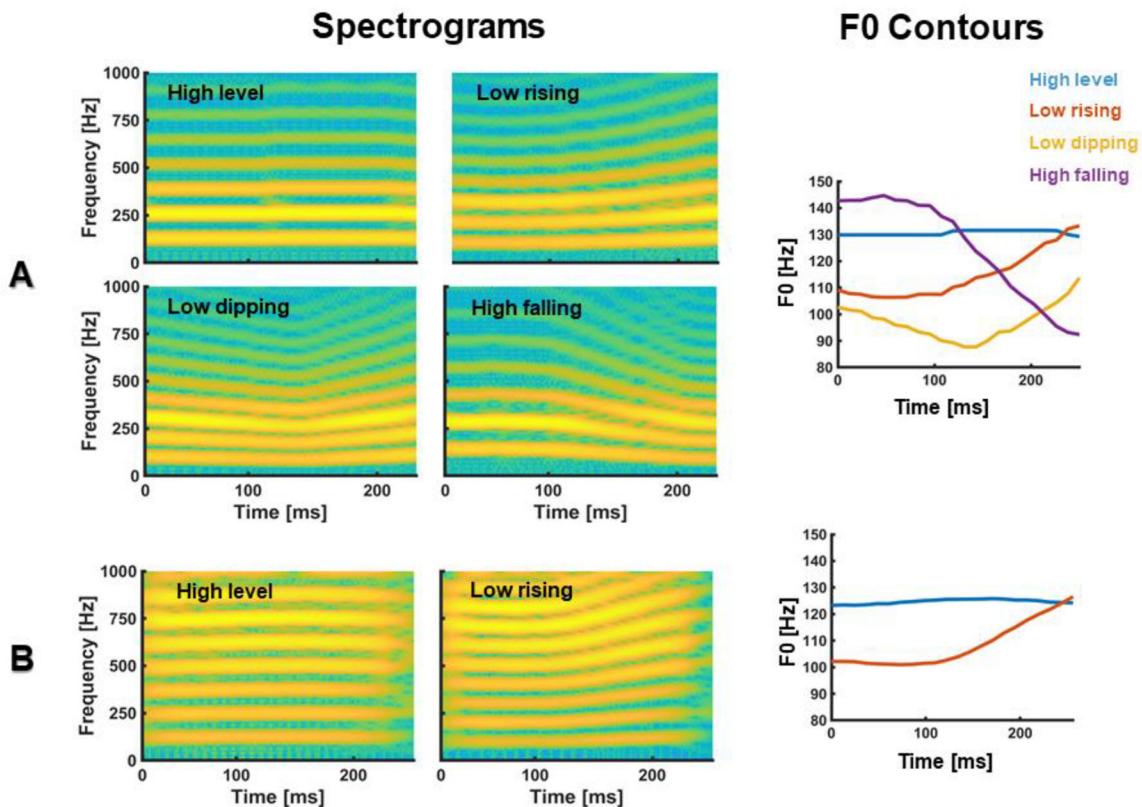


Figure 2. (A) Spectrograms and F0 contours of the waveforms of four prototypical Mandarin Chinese tones (from Reetzke et al., 2018). (B) Spectrograms and F0 contours of the two Mandarin tone waveforms used in our study.

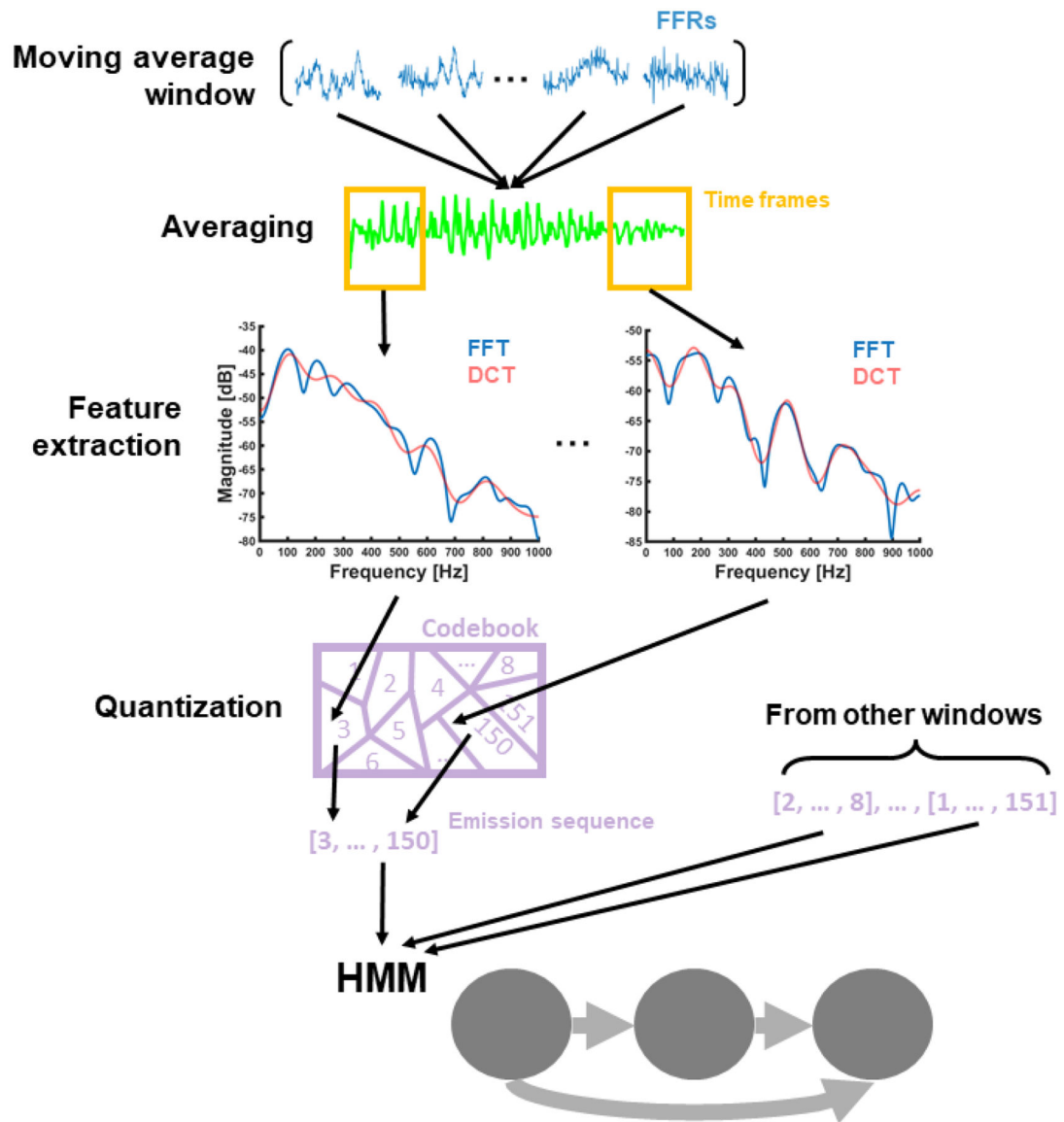


Figure 3. Schematic representation of the steps followed to generate the input of a generic HMM with no frequency band constraints. From top to bottom: FFRs within the same training or testing set were sub-averaged with a moving average window, and averaged responses were processed with the short-term Fourier transform. Fourier spectral frames were scaled in decibels and encoded into the first 13th coefficients of the discrete cosine transform (DCT). Each sequence of DCT frames (one for each averaged response) was quantized into a sequence of discrete HMM emissions. The HMM was trained and tested with the emission sequences obtained from the corresponding training and testing sets.

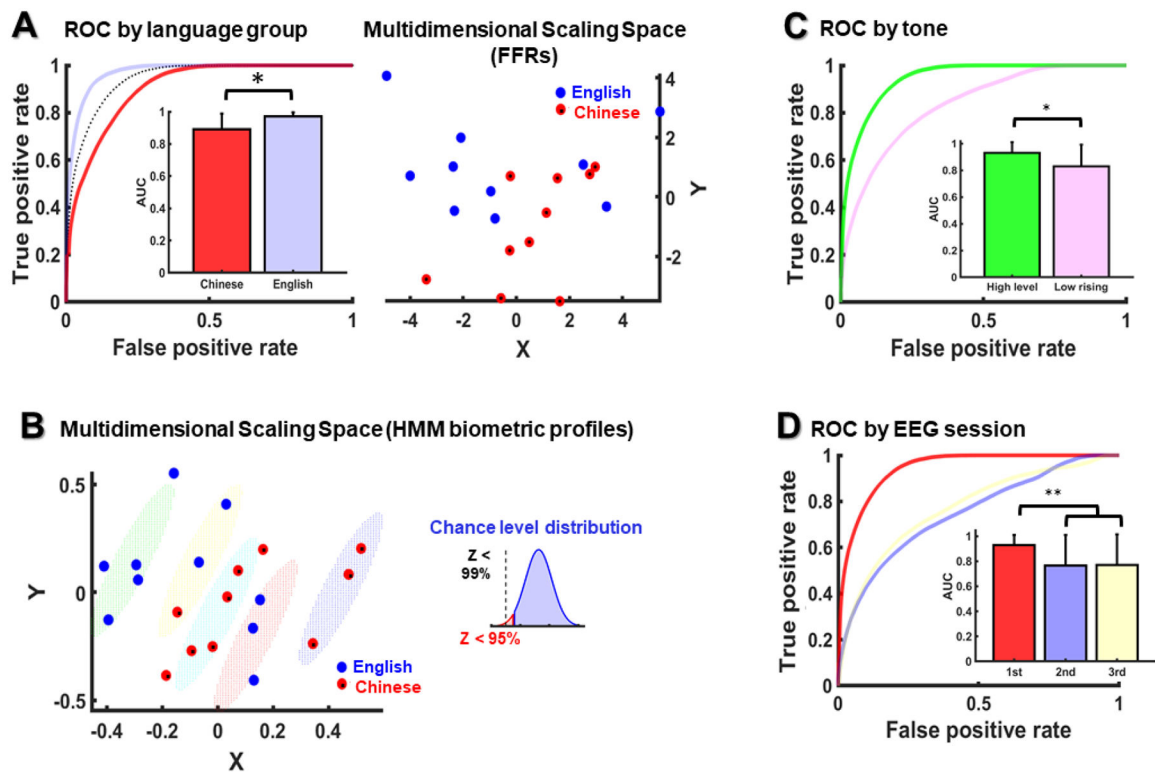


Figure 4.

(A) Mean AUROC for each language group (English and Chinese) for the same tone and EEG session (left panel), and English and Chinese averaged FFRs ($N=1000$) arranged by their relative Euclidean distance (right panel); in the left panel, the mean AUROC across all subjects is dotted in black. (B) English and Chinese biometric profiles clustered as a function of their relative biometric similarity. The small diagram in the top-right area of the panel shows the proportion of mean language dispersion values in the chance-level distribution that were larger than the mean dispersion value (Z) of the clusters shown below. (C) AUROCs of the HMMs trained with the high-level tone and tested with the low-rising and high-level tones. (D) AUROCs of the HMMs trained with the first EEG session and tested with the first, second and third sessions. Error bars are expressed in standard deviation units.

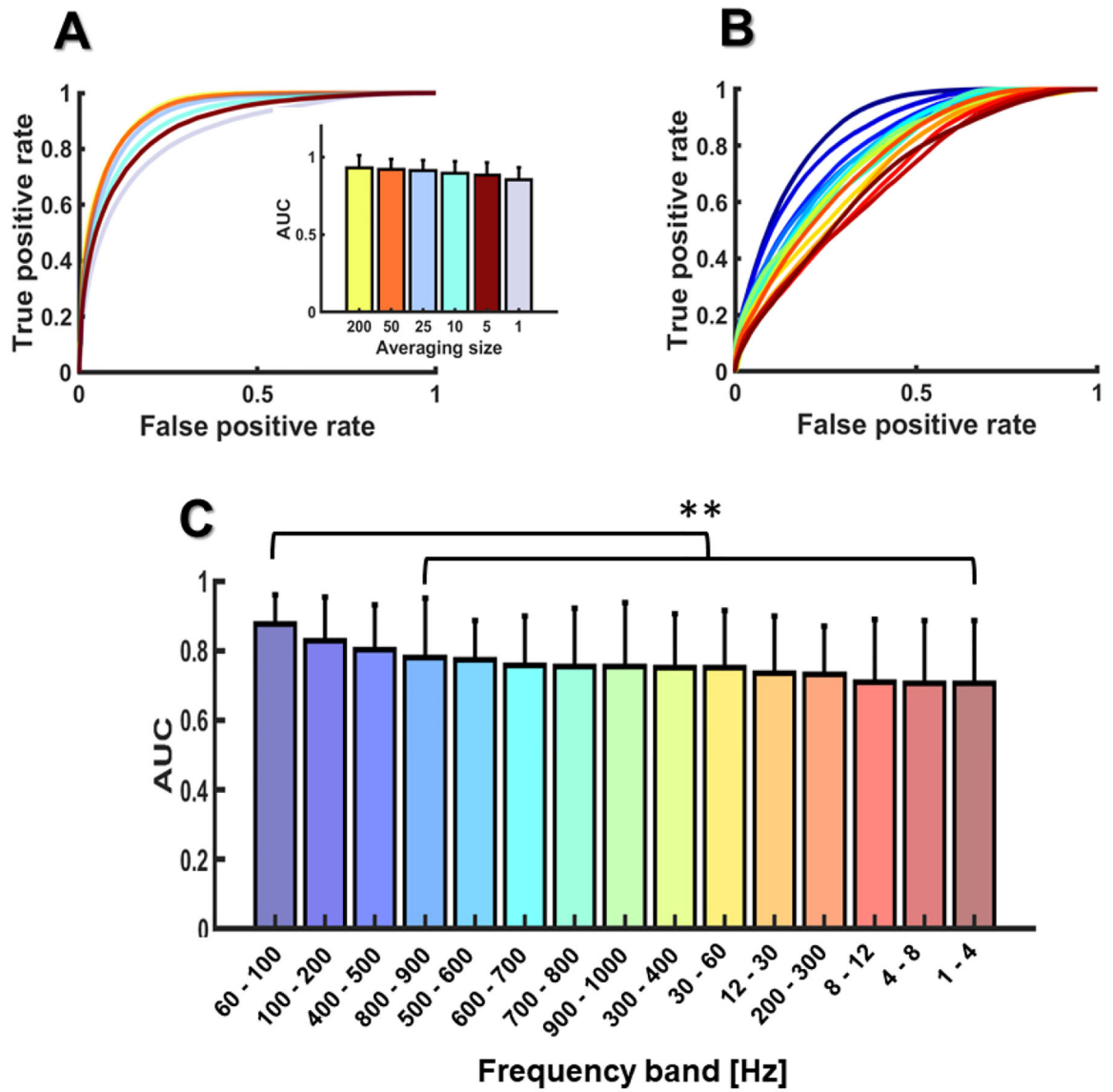


Figure 5. AUROCs as a function of the FFR averaging size (A) and the EEG frequency band (B, C). Error bars are expressed in standard deviation units.

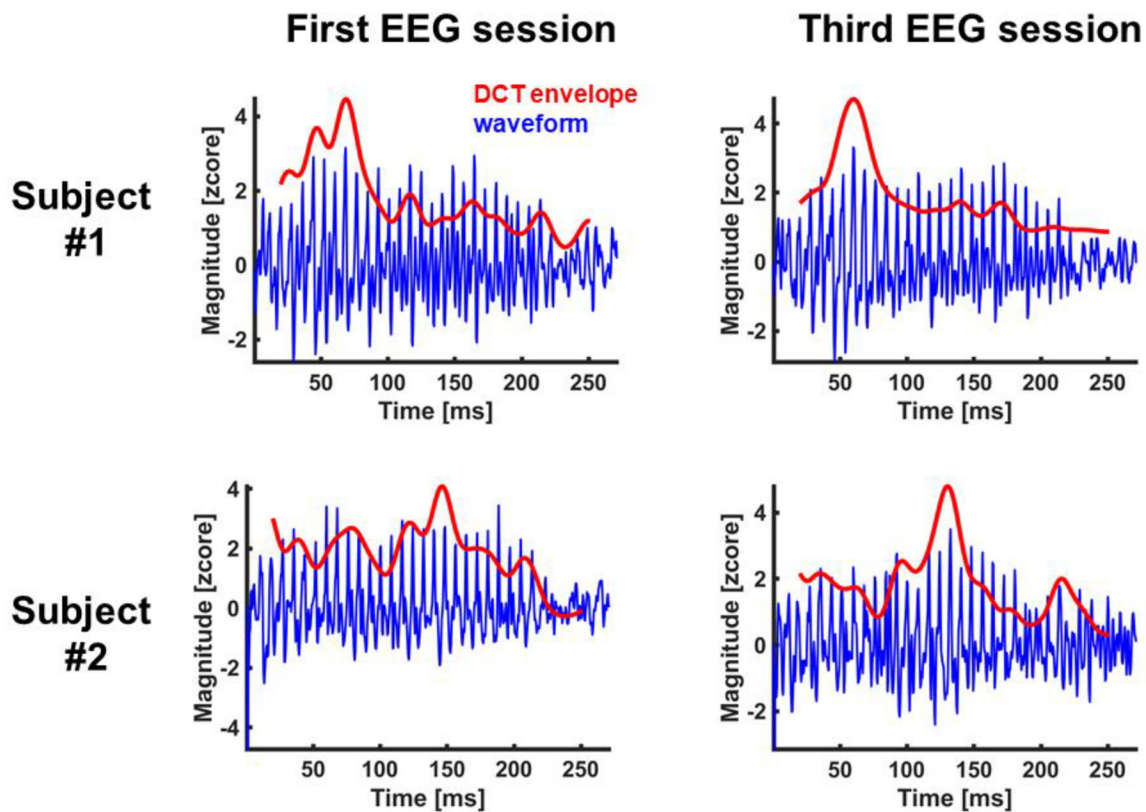


Figure 6.

FFR waveforms (waveform) from two listeners in the first and last EEG sessions. The figure also shows the FFR waveforms reconstructed from the first 13 coefficients of the discrete cosine transform of each spectrum slice provided by the short-term Fourier transform (DCT envelope). DCT spectral envelopes were projected back in the time domain using the inverse discrete Fourier transform.