

Amplitude-Binned Envelope Modeling for EEG Response Prediction in Naturalistic Soundscapes

Neurophysiology of everyday life group

Psychology Department

05 August - 30 November. 2024.

Student: Elnur Imamaliyev

Examiner / Supervisor: Dr. Martin Bleichner | University of Oldenburg

Supervisor: M.Sc. Thorge Haupt | University of Oldenburg

Table of Contents

Table of Contents.....	2
Abstract.....	3
1. Introduction.....	4
Study Objectives and Hypotheses.....	6
2. Materials and Methods.....	7
2.1.1. Dataset.....	7
2.1.2. Data Acquisition.....	7
2.1.3. Experimental Procedure and Conditions.....	7
2.2. EEG Preprocessing and Audio Extraction.....	9
2.3. Feature Generation.....	11
2.3.1. Standard envelope generation.....	11
2.3.2. Amplitude-binned (AB) envelope generation.....	12
2.3.3. Additional envelope models: Onset Envelope and Combined envelope modes...	13
2.4. TRF estimation.....	14
2.5. Study design and statistical analyses.....	15
3. Results.....	18
3.1. Choosing bin size parameters and methods - Search Grid.....	18
3.2. Sound Distribution.....	20
3.3. Channel selection - Topoplots.....	22
3.4. Prediction Accuracies: Condition averaged.....	23
3.5. Weights, Latencies, P2-N1 peaks (Group and Condition Averaged).....	24
3.6. Index Shuffled Binned-Envelope.....	26
3.7. Impact of Attention Conditions on AB Envelope Modeling.....	26
3.8. Combined Model Performance: Condition Averaged.....	28
4. Discussion.....	29
Bin Parameters and Binning Methods.....	29
Tracking Latency and Magnitude Changes.....	30
On Binned envelope model performance (H1).....	30
Combination of envelopes (H2).....	32
Future studies.....	32
5. Conclusion.....	33
References:.....	34

Abstract

Accurately predicting neural responses to naturalistic auditory stimuli remains a key challenge in auditory neuroscience. In this study, we used a amplitude-binned (AB) envelope model that partitions the continuous auditory envelope into discrete amplitude segments, thereby capturing the nonlinear aspects of neural processing. Electroencephalography (EEG) data were collected from participants immersed in a complex, real-world auditory environment during a 3D Tetris task, where concurrent speech, alarm, and beep sounds created variable attention demands. Using a multivariate Temporal Response Function (mTRF) framework, we optimized different binning strategies and normalization protocols to evaluate the performance of our AB envelope model against conventional Standard Envelope approaches.

Our results indicate that discretizing the auditory signal into amplitude bins markedly improves the prediction of EEG responses. Notably, when participants maintained focused attention, our AB model produced significantly higher prediction accuracies, especially in frontal and temporal cortical areas, and exhibited a robust linear association between stimulus intensity and neural response amplitude. In conditions where attention was more diffusely distributed, the distinction between model performances diminished, implying that selective focus enhances neural encoding of sound intensity variations. The careful calibration of bin parameters and normalization timing provides critical insights into the optimal design of neural encoding models. Furthermore, augmenting the AB model with an onset-based envelope to capture rapid auditory transitions further increased prediction accuracy by accounting for fast temporal dynamics absent in traditional methods. Statistical evaluation, including permutation tests and paired comparisons, confirmed the reliability of these improvements.

These findings shows the potential of multivariate envelope representations for improving neural response prediction in naturalistic settings and lay the groundwork for advanced neural encoding models applicable to real-world auditory processing.

1. Introduction

The human auditory system is a sophisticated biological mechanism capable of processing a wide range of acoustic stimuli. A fundamental question in auditory neuroscience is understanding the neural mechanisms underlying these perceptual abilities. While traditional studies often relied on simplified, artificial stimuli, recent research has increasingly turned to naturalistic soundscapes to better understand the brain's intricate processing of real-world auditory information.

Naturalistic stimuli, such as speech, music, and environmental sounds, offer significant advantages over artificial stimuli. They are ecologically valid, capturing the rich temporal and spectral characteristics of real-world sounds, and they engage a broader range of neural processes, including attention, memory, and emotion. By studying how the brain responds to naturalistic stimuli, researchers gain a more comprehensive understanding of auditory perception and its underlying neural mechanisms.

The use of naturalistic stimuli reflects real-world listening conditions more accurately, enabling the study of complex auditory environments compared to synthetic tones. These stimuli facilitate the investigation of dynamic features like speech modulation, noise masking, and communication signals, which are difficult to replicate using synthetic sounds. For example, in speech neuroscience, naturalistic stimuli provide a more realistic representation of how the brain processes sounds in everyday settings, uncovering neural responses that simplified stimuli might overlook (Theunissen & Elie, 2014). The use of naturalistic soundscapes, such as human speech or animal vocalizations, also allows researchers to model real-world listening conditions (Theunissen et al., 2001; Machens et al., 2004). Recent research highlights the benefits of this approach in uncovering how auditory perception, attention, and communication operate in daily environments, making findings more relevant and applicable to real-world scenarios.

To study the neural responses to acoustic stimuli linear models have been widely employed. These models assume that the brain's response to a stimulus is a linear function of its acoustic properties. While successful in explaining certain aspects of auditory processing, linear models face limitations in capturing the complex, nonlinear nature of neural responses.

One effective approach to understanding sensory processing involves models that link sensory inputs to neural responses (Wu et al., 2006). These models predict neural activity responses to novel stimuli, which is crucial for uncovering sensory processing mechanisms (Carandini et al., 2005). Features like the envelope, which represents the slowly varying amplitude fluctuations of a sound signal over time, and the spectrogram,

a visual representation that displays the frequency components of a sound signal across time, are commonly used to predict neural responses in both animals and humans (David et al., 2007; Lalor & Foxe, 2010). The envelope provides information about the overall intensity dynamics of a signal, while the spectrogram captures detailed spectral and temporal patterns, aiding in understanding how the brain processes complex auditory stimuli.

Traditionally, linear receptive field models describe neural responses as a weighted sum of stimulus features, such as contrast in vision or amplitude in auditory signals. While effective in early sensory systems (Boynton et al., 1996), these models often oversimplify neural dynamics by assuming consistent scaling of responses with stimulus intensity (Crosse et al., 2016). In auditory neuroscience, linear models often use spectrotemporal receptive fields (STRFs) to characterize neural responses to sound stimuli with broad spectrotemporal statistics (Aertsen & Johannesma, 1981).

One popular method for linking neural responses to continuous auditory stimuli is the Temporal Response Function (TRF). TRFs describe how EEG signals respond to dynamic auditory features, assuming that stimulus intensity affects only the magnitude of the neural response, not its timing or morphology (Lalor et al., 2009). However, this assumption is limited, as response latency and shape often vary with stimulus amplitude (Beagley & Knight, 1967). For instance, TRFs can incorporate one-dimensional inputs like amplitude envelopes or multivariate features such as spectrograms, providing insights into how the brain processes auditory stimuli in real time.

Recent studies demonstrate that the brain often exhibits nonlinear responses to acoustic stimuli. For instance, Drennan et al. (2019) showed that binning the envelope into different amplitude bands captures a significant portion of neural variability. This approach accounts for nonlinearities by representing the envelope as discrete amplitude bins, providing a more accurate depiction of the neural responses to complex auditory stimuli.

While these findings offer valuable insights into speech processing, their generalizability to more complex soundscapes remains unclear. Rosenkrantz et al. (2023) investigated neural responses to various naturalistic soundscape environments. Their study highlights the diverse neural mechanisms the brain employs to process complex auditory scenes. By incorporating the absolute value of the Hilbert transform of audio signals, followed by low-pass filtering and resampling, their TRF models effectively captured dynamic sound features, enhancing the prediction of neural responses in naturalistic environments.

Additionally, research suggests that altering the way sound features are modeled significantly impacts EEG response prediction accuracy. Haupt et al. (2024)

demonstrated that combining envelope models, such as the envelope and onset, can predict neural responses in naturalistic auditory conditions. Their findings emphasize the importance of adapting neural models to account for the complexities of naturalistic auditory environments. While no single feature set is optimal, these studies underscore the need for models that reflect the intricate neural dynamics associated with real-world listening scenarios.

In summary, naturalistic stimuli and advanced modeling techniques, such as TRFs and amplitude binning, provide valuable insights into the complex neural mechanisms underlying auditory perception. These approaches improve our understanding of how the brain processes dynamic and ecologically valid soundscapes, enhancing the relevance of auditory neuroscience findings to everyday listening conditions.

Study Objectives and Hypotheses

In this study, we aim to assess the AB envelope model and extend the findings of Drennan et al. (2019) on EEG predictions by testing their method on a naturalistic soundscape. To apply this method in a naturalistic soundscape, we have used an audio-motor task dataset based on the work of Rosenkranz et al. (2023), which includes speech and non-speech sounds. This complex auditory environment provides a realistic framework for evaluating whether the brain responds non-linearly to different intensities in a dynamic soundscape.

Our first hypothesis posits that dividing the auditory envelope into multiple amplitude-based bins, each capturing specific levels of stimulus intensity, will enhance the accuracy of response predictions compared to using a standard envelope.

Our second hypothesis suggests that combining the envelope models with different features, such as utilizing both the onset envelope (emphasizing amplitude increases) and the AB envelope, may further improve prediction accuracy while emphasizing these changes.

To test these hypotheses, we will investigate whether dividing the stimulus envelope into amplitude-based segments and applying a multivariate TRF to each segment can outperform the traditional audio features and reveal how stimulus intensity affects response morphology and latency. We will also assess the relevance of different features, such as stimulus onset and envelope, while comparing them to each other, AB envelope, etc., in predicting neural responses. This research could provide valuable insights into the influence of sound feature modeling on EEG predictions and expand our understanding of neural processing in BTL environments.

2. Materials and Methods

2.1.1. Dataset

This study uses EEG and audio data from an experiment examining attentional focus in complex audio-visual environments, specifically designed to simulate workplace settings like an operating room. The auditory stimuli were crafted to blend background noise with both relevant and irrelevant sounds, offering rich insights into how different attentional focuses influence performance in dynamic environments. The dataset features 20 participants (14 females, aged 20–30) who performed two conditions: (1) narrow-focus (attending to specific auditory targets) and (2) wide-focus (beep tone was coming from multiple directions and embedded into other tones). Participants engaged in a 3D Tetris task while exposed to various tones, including background noise, task-relevant and task-irrelevant sounds, and speech. The data can be accessed via Zenodo (www.zenodo.org/record/7147701, Rosenkranz et al., 2023).

2.1.2. Data Acquisition

The EEG data was captured using a wireless 24-channel amplifier (SMARTING, mBrainTrain, Belgrade, Serbia) connected to the back of the EEG cap (EasyCap GmbH, Hersching, Germany) with Ag/AgCl passive electrodes. The reference and ground electrodes were positioned at Fz and AFz, respectively. The data was recorded at a sampling rate of 500 Hz and the impedance of electrodes was kept below 20Ω before the recording. The audio presented during the experiment was sampled at 44.1 kHz. The data is transmitted via Bluetooth from the amplifier to a Bluetooth dongle (BlueSoleil) connected to a computer (Dell Optiplex 5070). To allow for the joint analysis of auditory stimuli and neural responses, all data streams recorded were synchronised using the Lab Recorder software, which is based on the Lab Streaming Layer. All participants were informed about the procedure and had to sign the informed consent before the EEG measurements.

2.1.3. Experimental Procedure and Conditions

In the original study from Rosenkranz et al., 2023, participants performed a complex audio-visual-motor task - a 3D Tetris game, an adapted version of the original Tetris game developed by Kalarus (2021). The game involved placing three-dimensional blocks to form and remove layers for points. Participants controlled block rotation with the left hand and positioning with the right. Unlike classic Tetris, the game restarted at the bottom layer upon stacking too high, allowing for continuous gameplay but resulting in point loss.

While playing, participants listened to a complex soundscape comprising continuous hospital background noise and five discrete stimuli: task-irrelevant speech (from left behind, -135°), two irrelevant sounds (from the left, -90°), task-relevant instructions of where to place specific blocks (from the front, 0°), an alarm (45°), and a beep (the same direction as the other stimuli). Spatial separation was achieved via the Head-Related Impulse function (Kayser et al., 2009).

The study mainly focused on attention and therefore had two attention conditions: Narrow and Wide. The task required attending to speech instructions guiding block placement and, depending on the condition, either the alarm (narrow condition) or the alarm and the beep (wide condition).

1. **Narrow Condition:** Participants were instructed to additionally attend to the alarm, i.e., participants had to attend to the task-relevant speech and the alarm. The alarm was long, had high intensity, and was always presented from the same direction, thus it was not necessary to attend to the rest of the soundscape to detect it.
2. **Wide Condition:** Participants were instructed to attend to the beep, i.e., participants had to attend to the task-relevant speech and the beep. The beep was short, had low intensity, and was integrated into other stimuli, thus the whole soundscape had to be monitored to detect it.

Note that the soundscape was conceptually the same for both conditions. To summarize, the difference between the two conditions was the instruction on which target should be attended to. The targets of the narrow and wide conditions were the alarm and the beep, respectively.

Participants pressed the space bar upon hitting the targets, earning points for correct responses and following the speech instructions, while points were deducted for not following instructions. They completed two 18-minute games under narrow and wide conditions, receiving different instructions for each condition.

To familiarize themselves with the game, participants received written instructions. They then underwent general training without auditory stimuli and training for the relevant speech instructions. Before each condition, participants also performed condition-specific training, receiving feedback on whether they correctly detected the target. EEG was not recorded during the training games. Before the start of each condition's game, resting EEG was recorded by instructing participants to first focus on a fixation cross and then close their eyes for 1 minute each.

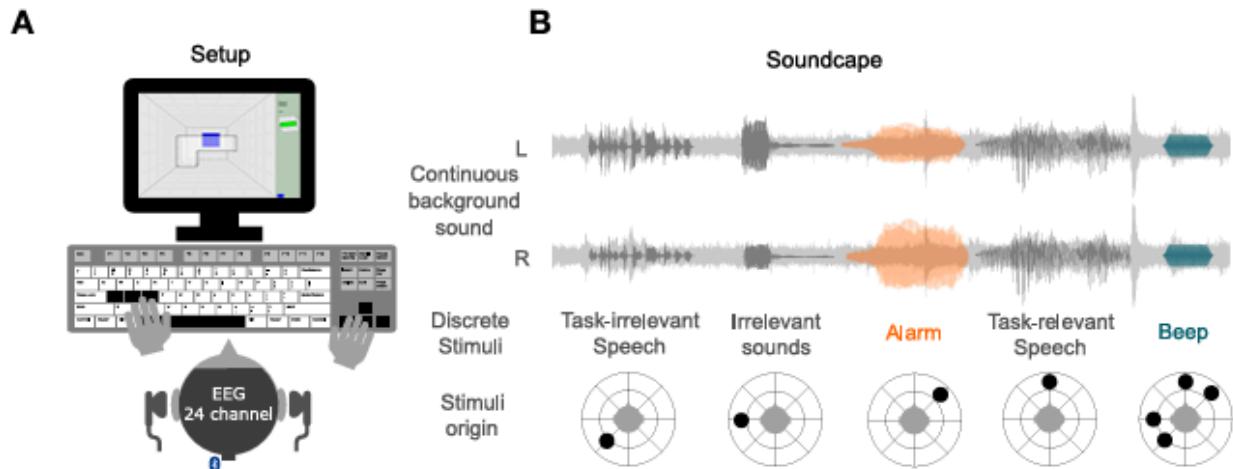


Figure 2.1. Taken from Rosenkranz et al. (2023). (A) **Experimental Setup:** Participants played 3D Tetris (with their left hand the participant controlled the rotation of a block, with their right hand the position of a block). The soundscape was presented via headphones. EEG was recorded using a 24-channel mobile EEG setup. (B) **Soundscape:** A continuous background sound was presented throughout the task. Discrete stimuli were subsequently presented. The alarm was the target in the narrow condition, while the beep was the target in the wide condition. The alarm was presented from one direction, while the beep was presented from any direction as the other sounds. If participants detected a target, they should press the space bar.

For more details, refer to Rosenkranz et al., 2023.

2.2. EEG Preprocessing and Audio Extraction

2.2.1. EEG extraction and preprocessing

The EEG data was preprocessed in MATLAB (version 2024b, MathWorks, Natick, MA) using the EEGLAB plugin (v2024.1, Delorme and Makeig, 2004), mTRF toolboxes (Crosse et al., 2016) and custom scripts. For each experimental condition (narrow and wide), and for each participant, we did the extraction separately. Initially, the EEG data were extracted from the XDF format, and then event latencies (labels) were used to synchronise the EEG with the audio events and saved as .set format.

After extraction from the .set files, ICA weights are derived. First, the EEG datasets for each condition are loaded using *pop_loadset*, and channel locations from the dataset are stored for both conditions and they are merged using *pop_mergeset* to ensure ICA weights are computed over both conditions as the assumption was that the artifact locations are similar between conditions. The data is then downsampled to 250 Hz using the *pop_resample* function, and converted to double precision using *double*

function, which is required for ICA. High-pass and low-pass filters are applied using *pop_firws* with cutoff frequencies of 1 Hz and 42 Hz, respectively, and a Hann window. Specifically, Highpass ('fcutoff', 1, 'ftype', 'highpass', 'wtype', 'hann', 'forder', 568); lowpass, ('fcutoff', 42, 'ftype', 'lowpass', 'wtype', 'hann', 'forder', 128); Channels with poor signal quality are removed using the *clean_channels* function, and the number of removed channels is recorded. The data is epoched using *eeg_regePOCHS* with a recurrence of 1 second, and epochs with artifacts are removed using *pop_jointprob* with a threshold of 3 standard deviations. ICA weights are computed using *pop_runica* with the 'runica' algorithm, extended mode, and concatenation of data segments. ('icatype', 'runica', 'extended', 1, 'interrupt', 'on', 'concatenate', 'on'); ICA weights, along with the sphere and channel indices, are saved. Each raw dataset is then reloaded, bad channels are removed again, and the saved ICA weights are added. The IC components are labeled using *pop_iclabel*, and the datasets with ICA weights and labels are saved. This comprehensive process ensures that excessive artifacts are removed and the data is ready for further preprocessing.

The preprocessing continues as EEGLAB is initialized without the graphical user interface using *eeglab('nogui')*, which is suitable for batch processing. The EEG dataset is then loaded using *pop_loadset*. Components labeled with a high probability of being artifacts are rejected using *pop_icflag* with the specified probability ranges [0.7 1; 0.7 1; 0.6 1; 0.7 1; 0.7 1], corresponding to muscle, eye, heart, line noise, channel noise artifacts and the dataset is stored using *eeg_store*. The identified components are removed using *pop_subcomp*, and the dataset is stored again. A low-pass filter is applied using *pop_firws* with a cutoff frequency specified by */p*, a rectangular window type, and a filter order of 100. The data is then downsampled to 100 Hz using *pop_resample* to save processing time. A high-pass filter is applied using *pop_firws* with a cutoff frequency of 0.3 Hz, a Hann window type, and a filter order of 518. Missing channels are interpolated using *pop_interp* with spherical interpolation based on the original channel locations (*EEG.urchanlocs*). The data is re-referenced to the mastoid electrodes (TP9 and TP10) using *pop_reref*, with the option to keep the reference channels. Finally, the mastoid channels are removed using *pop_select* as they no longer hold meaningful information after re-referencing. These comprehensive steps ensure that the EEG data is thoroughly preprocessed, removing artifacts and preparing it for further analysis.

2.2.2. Audio extraction

Initially, the Audio data were extracted from the XDF format. For the audio processing, the audio stream is loaded using *pop_loadxdf* with the specified file path and stream name 'TetrisAudio'. Then event latencies (labels) were used to synchronize the EEG with the audio events. The audio data, initially stored as int16, is converted to double

precision. The stereo audio channels are averaged to create a mono signal, which is then scaled by dividing by 3276.7. This processed audio data is stored back into the *audio_struct* structure. The epoch duration for the audio data is calculated by finding the latency difference between the '*game_end*' and the task-specific '*game_start*' events, divided by the audio sampling rate. The audio data is then epoched using *pop_epoch* based on the '*game_start*' event and the calculated epoch duration. The processed audio data and the *audio_struct* structure are saved to .mat files with filenames that include the subject and task identifiers. Then the audio data structures for the narrow and wide game conditions are loaded. The loaded audio data is then written to WAV files using audiowrite, with the narrow game audio and wide game audio saved separately.

2.3. Feature Generation

For this study we used Normalized Standard envelope, AB envelope, Onset envelope and Combined features: AB envelope + Onset envelope;

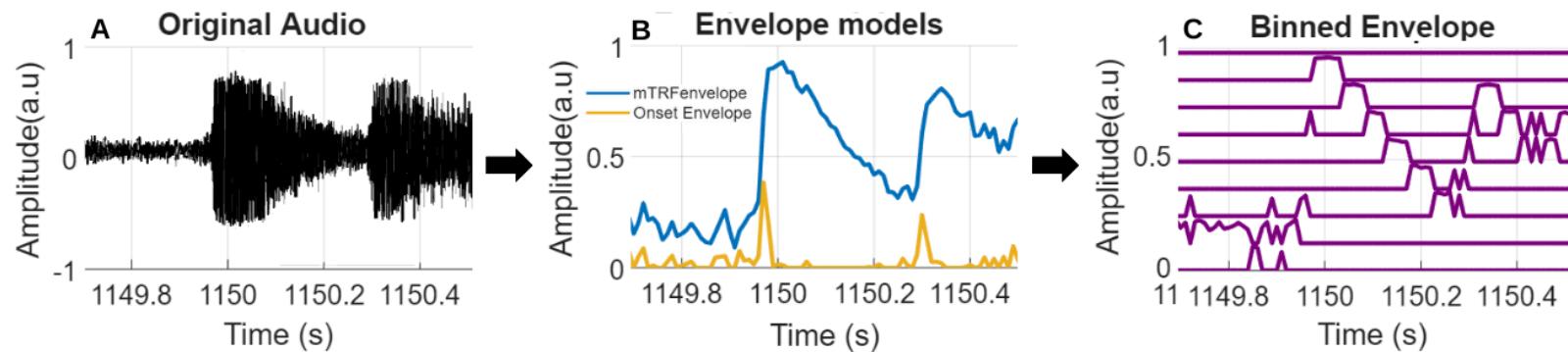


Figure 2.3. Envelope models. Figure shows how amplitude changes in the audio sample relate to the computed envelope models. A - The original audio sample waveform. Generated envelope models: B - Univariate: Standard and Onset envelopes. C - Multivariate: Amplitude-Binned (AB) envelope.

2.3.1. Standard envelope generation

Previous studies have demonstrated that EEG-measured neural activity can track the acoustic envelope of auditory stimuli (Drennan and Lalor, 2019; Giraud and Poeppel, 2012; Luo and Poeppel, 2007). We decided to derive the envelope, which represents the sound's amplitude over time. Here, we specifically decided to use the mTRFenvelope function, from the mTRF toolbox (Crosse et al., 2016) as it has been shown to be suitable to predict neural data for this specific data set (Haupt et al., 2024).

The mTRFenvelope function estimates the temporal envelope of an audio signal by resampling and averaging the signal's intensity. It begins by taking the input audio signal (x) and resampling it from the input sampling frequency $fsin$ to the output sampling frequency $fsout$. This is done by averaging the square of the nearest neighbors every $fsin/fsout$ samples, then taking the square root to compute the root mean square (RMS) intensity. The RMS values are logarithmically scaled to model human auditory perception, using a default compression factor of $\log_{10}(2)$. The function also allows for specifying a window size for averaging, which can smooth the envelope by overlapping data used to estimate adjacent output frames. Additionally, a buffer of initial data can be concatenated to the beginning of the signal to center the first window at time $t=0$. The function returns the estimated temporal envelope y , the corresponding time vector t , and a cache of intermediate values. This process helps in creating a smooth and stable representation of the audio signal's envelope, useful for various auditory processing tasks.

After generation, envelopes are normalized to be between 0 and 1 range, using the *normalize* function. The result of normalisation is the Normalized Standard Envelope feature, which we refer to as the Standard Envelope model. (Figure 2.2A)

2.3.2. Amplitude-binned (AB) envelope generation

Drennan and Lalor (2019) explain their chosen empirical bin size after comparing the prediction accuracies attained across a range of bin sizes. Broader bins perhaps being less able to capture changes in the response with amplitude, and narrower bins perhaps suffering from the limited amount of data available for training. In their study, from 8 to 80 dB range for the Amplitude Binned Envelope (ABenvelope) is used. We slightly modified their process by choosing the binning range and number of bins, with the step size of the dB values calculated automatically. Specifically, the bounds were defined first, and then the step size was computed based on bin numbers. To optimize the AB envelope model for a significant improvement over the Standard Envelope model, we conducted a search grid analysis across selected parameter ranges. For the lower bound parameters, we tested values ranging from 0, 6, 10, 20, 30, 40, 50, 60, and 70 dB, while for the upper bound parameters, we used values ranging from 6, 10, 20, 30, 40, 50, 60, 70, and 80 dB systematically. We also tested varying numbers of bins: 4, 5, 6, 8, 9, 12, and 16, with the condition that the upper bound value must be greater than the lower bound. To evaluate model performance of each parameter combination, we applied the Wilcoxon signed-rank test ($\alpha = 0.05$) to compare correlation values between the standard envelope and the AB envelope, with all results saved for systematic comparison.

The AB envelope is derived through a systematic process that begins by defining the parameters for binning, including the lower bound and upper bound bin values (0 and 0.98, respectively) and the number of bins (8). These bin parameters were chosen based on a search grid analysis results, normalized from 0 to 62.38dB.

The bin edges are calculated using *linspace* to create equally spaced divisions within this range. To work with amplitude values instead of dB, the bin edges defined in dB are converted to a linear scale using the formula $10^{(\text{dBbinEdges} / 20)}$, and then normalizing the resulting range to between 0 and 1.

Next, the envelope generated by the mTRFenvelope function is normalized to the 0–1 range. The histcounts function is used to compute the indices of normalized envelope values within each defined bin to store the binned envelope values, based on the computed indices. Then an empty matrix is initialized to store the binned envelope values. Then a loop populates this matrix by assigning the normalized envelope values to their corresponding bins based on the computed indices. As a result, a structured representation of the envelope's distribution across the defined amplitude ranges is generated. Finally, the binned envelope values are normalized again (Figure 2.2C.) This final normalization ensures the stability of the resulting TRF, as it mitigates the effects of large amplitude fluctuations that could otherwise skew the model's predictions.

2.3.3. Additional envelope models: Onset Envelope and Combined envelope modes

Additionally, Drennan et al. (2019) suggest that combining the AB envelope with the onset envelope might increase prediction accuracy, as these models capture complementary information on envelope tracking. To test this idea, we employed both the onset envelope and a combination of the onset envelope with the AB envelope to assess whether this combination would lead to an increased correlation, in comparison to using AB envelope alone.

To generate the onset envelope, the difference between consecutive normalized envelope values was calculated to detect changes. Only positive differences were retained, with negative values set to zero, to highlight the onset of sounds. The onset points were then detected using a specified threshold and minimum peak distance, resulting in the onset envelope. The onset envelope was then normalized.

The combined envelope was generated by concatenating the onset envelope on top of the AB envelope, resulting in the AB envelope + onset envelope.

2.4. TRF estimation

We used **Temporal Response Function (TRF)** estimation to investigate how derived stimulus features are represented in the neural signal (EEG response). TRF estimation links stimulus features to EEG responses, implemented using the mTRF Toolbox (Crosse et al., 2016). The method assumes that the EEG signal, $y(t)$, is modeled as the convolution of an input stimulus feature, $x(t)$, with an unknown system response, $w(\tau)$ (the TRF), plus noise:

$$y(t) = w(\tau) * x(t) + \text{noise}$$

where τ represents the range of time lags over which the TRF is estimated. The TRF, $w(\tau)$, is derived by minimizing the mean squared error between the predicted and recorded EEG signals using ordinary least squares (OLS). Ridge regression is applied for regularization during the normalization process to prevent overfitting (Lalor et al., 2009).

The time lags range from -100 ms to 400 ms to capture both pre- and post-stimulus neural responses. The neural response, $r(t,c)$, is modelled as the convolution of stimulus features with channel-specific weights over time lags:

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

where c represents the EEG channel, t represents time, and τ is the time lag. The TRF weights are computed using the formula:

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

where S is the design matrix containing the lagged stimulus features, r is the EEG response, and λ is the regularization parameter optimized via cross-validation. Zero-padding is applied to the stimulus matrix at non-zero lags to ensure causality (Mesgarani et al., 2009).

The dimensionality of the design matrix is determined by the number of features (N_f), lags (τ) and time points (T): $T \times (N_f \cdot \tau)$. For example, the AB envelope results in a design matrix of dimensions $T \times (8 \cdot 51)$, while the Standard envelope results in $T \times 51$.

The TRF weights are analyzed to interpret both their morphology (temporal patterns across time lags) and topography (spatial distribution across EEG channels). The derived weights reflect the similarity between stimulus features and neural responses, with the inner product in the design matrix capturing this relationship.

In our study, the TRF model used the following stimulus features as input: the Standard envelope (univariate), the Amplitude Binned (AB) envelope (multivariate), the Onset envelope (univariate), and a Combined envelope (AB + Onset) (multivariate). The Standard envelope and Onset Envelope results in weights of dimensions $1 \times 51 \times 22$ (1 feature, 51 lags, 22 channels), while the AB envelope generates weights of $8 \times 51 \times 22$ (8 bins, 51 lags, 22 channels). Similarly, the combined envelope generates weights of $9 \times 51 \times 22$ (8 bins + 1 feature, 51 lags, 22 channels)

2.5. Study design and statistical analyses

The data analysis was performed using MATLAB (version R2024b, MathWorks, Natick, MA) with the additional custom scripts. Part of the analyses were inspired by the work of Drennan et al. (2019) and Thorge et al. (2024). In this study, we analyzed EEG data from multiple subjects performing 3D Tetris peace replacement tasks, under narrow and wide attention condition to investigate the relationship between neural responses and auditory stimuli. For each subject-task combination, EEG and audio data were loaded using a custom function, returning the EEG signal, audio data, and associated sampling frequencies.

Training and Feature Extraction

Four features were generated to extract relevant features from the audio signals: standard envelope, AB envelope, Onset Envelope, and combined envelope models. Then the predicted EEG response was split accordingly to ensure comparability between modelled neural data and audio. These four feature sets were then used for modelling. For each feature set, the data were partitioned into training and testing segments using cross-validation, with 10-fold cross-validation for training and 1-fold testing. The analysis was conducted within a time window of -100 ms to 400 ms. The optimal regularization parameter was determined by cross-validation, testing values of lambdas spanning from 10^{-4} to 10^4 , with 10 equally spaced values. The optimal lambda was selected based on the mean model performance across folds. Importantly, the parameter optimization was done separately for each stimulus representation and subject, so that we were left comparing each model based on its respective optimal performance.

Using the chosen optimal regularization parameter, the model was trained and evaluated on unseen test data. The performance of each model was assessed by quantifying how accurately the predicted neural response matched the actual recorded response using Pearson's correlation coefficient.

Statistical methods

The normality of the correlation values for each model was verified through the Anderson-Darling test. Model comparisons were performed using paired-sample t-tests and Wilcoxon signed-rank tests, with a significance threshold set at $\alpha = 0.05$ for both. For t-tests, Cohen's d effect size was computed by dividing the t-value by the square root of the sample size, providing a measure of the magnitude of the difference between models.

Permutation Test and Bin Index Shuffling

Permutation tests were also employed to assess the null distributions of the envelope models. For the Naturalistic dataset, circularly random-shifted envelopes were generated and used in the cross-validation procedure. This process was repeated 100 times to establish the null distribution of each envelope model for every subject. By comparing the observed performance of the models with the null distribution, the significance of the model's predictive power could be determined, accounting for the variability inherent in the data. This approach further strengthened the validity of the findings and ensured that the observed results were not due to random fluctuations.

To test the specificity of the AB envelope model, bin indices were randomly shuffled (e.g., swapping the 1st bin with the 5th) to disrupt the relationship between amplitude levels and neural responses. The shuffled envelopes were used in the same 10-fold cross-validation framework as the original analysis. Prediction accuracies were compared using paired-sample t-tests and effect sizes (Cohen's d). This analysis validated the AB envelope's relevance by ensuring its performance was tied to meaningful neural representations.

Analysis of Weights and Latencies

The TRF models generated using the AB envelope allowed us to analyze weights and latencies for each of the eight amplitude bins independently. Each weight vector captured the temporal response (8 bins \times 51 lags \times channels) for a specific amplitude range, producing a multivariate representation of neural dynamics across bins. To simplify the analysis, channel-specific weights were averaged across channels, resulting in a 51-value row vector for each bin (8 bins \times 51 lags) that represents the averaged temporal response over lags.

To characterize the TRF weights, we focused on the N1 and P2 peak components, analyzing their magnitude and latency. These components are widely recognized in auditory processing and event-related potential (ERP) literature as markers of sensory and cognitive processing. Drennan et al. (2019) analyzed P1 and N1 peaks within

predefined latency ranges of 0–130 ms (P1) and 70–210 ms (N1). However, in our dataset, the P1 component was inconsistent and less pronounced, prompting us to focus on the P2 component instead. Initially, we analyzed P2 within a range of 70–200 ms but observed variability that required further refinement. To ensure robust and literature-informed peak detection, we reviewed prior studies, including Perrault and Picton (1984), which suggested latency ranges of 40–160 ms for N1 and 150–275 ms for P2.

Based on these guidelines, we selected adjusted latency ranges of 20–150 ms for N1 and 70–200 ms for P2. These ranges were chosen after visually inspecting weight plots across subjects and refining the windows to align with the most pronounced peaks observed in the data.

3. Results

3.1. Choosing bin size parameters and methods - Search Grid

The binning of the envelope depends on three parameters, the bin size, the lower and upper bound. The choice of bin size parameters and methods for the AB envelopes has a crucial effect on correlation accuracy. Bin size parameters were empirically chosen after evaluating prediction accuracies across various sizes, balancing the need for sufficient detail against the limitations of available training data.

We ultimately applied a logarithmic binning strategy, as described in Drennan et al. (2019), with an added normalization step before applying the conversion formula. (Table 3.1.2.) This strategy aimed to optimize the correlation performance of the AB envelope model.

For both narrow- and wide-attention conditions, we conducted a search grid analysis across selected parameter ranges and strategies. The lower bin edges were defined from 0 to 70 dB, the upper bin edges from 0 to 80 dB, and the number of bins tested varied between 4 and 16. The following conversion formula was applied:

$$10^{(\text{dBbinEdges} / 20)}$$

normalized the output, and fed it into the TRF function. To compare the performance of the standard and AB envelopes, we performed a Wilcoxon signed-rank test ($\alpha = 0.05$), categorizing the bin size parameters into three groups: significantly higher, significantly lower, or not significant, based on the results.

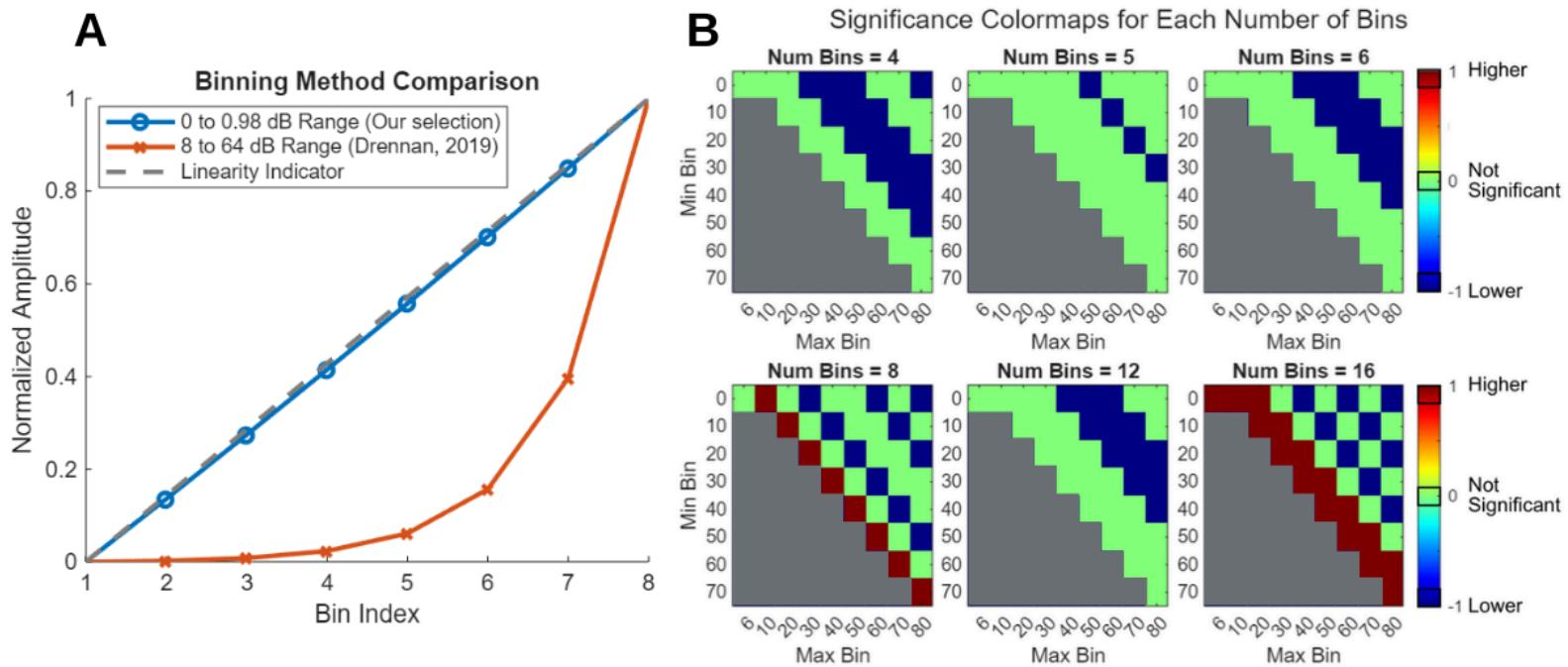
For the narrow-attention condition, significant improvements were observed with bin sizes of 8 and 16, especially when there was a 10–20 dB difference between the lower and upper bounds. This pattern formed a diagonal trend in the search grid, which was primarily attributed to the normalization effects (Figure 3.1.1B). Normalization emerged as a key factor in achieving parameter equivalence. For instance, configurations such as 0–10 dB (8 bins) and 10–20 dB (8 bins) produced identical normalized values, demonstrating that effective normalization is more critical than the absolute bounds themselves.

To further optimize, we applied a second approach where we used normalized values for the dB bin edges but ensured that the initial bin edges were normalized between 0 and 1 before applying the logarithmic conversion formula. This adjustment significantly improved the prediction accuracies of the AB envelope model. Further refinements in the normalized range showed that the 0–0.98 (8 bins) configuration was particularly effective, producing results similar to the 0–62.72 dB (8 bins) range.

In contrast, when this approach was applied to the wide-attention condition, no optimal parameters were identified that improved the AB envelope's performance over the standard envelope.

To compare our binning strategy with that used in Drennan et al. (2019), we first identified the optimal values that enhanced the performance of the AB envelope in the narrow-attention condition. We defined the bin parameters with a lower bound of 0, an upper bound of 64, and 8 bins. Two approaches were then applied: the data was first normalized using our study's version, and a non-normalized approach (as described in Drennan et al., 2019) was also applied before the conversion formula was implemented. After applying the conversion formula, we normalized the converted values and plotted them to visualize the bin edge parameters (Figure 3.1.1A). The results demonstrated that normalization prior to the conversion step resulted in post-conversion values that more closely followed a linear pattern.

Figure 3.1.1: Binning and Search Grid Performance Comparison. (A) Logarithmic binning with and without the normalization step. The figure compares our binning method with the binning method from Drennan et al. (2019). **(B) Image plots of the Search Grid on the Narrow-Attention Condition.** Gray values represent non-calculated range, blue values indicate significantly lower performance, red values highlight significantly higher performance, green values represent non-significant results compared to the standard envelope model.



Parameters	Defined Bin Edges	First Normalization	Logarithmic Conversion	After Second Normalization
0 to 62.72, 8 bins	[0, 7.84, 15.68, 23.52, 31.36, 39.20, 47.04, 54.88, 62.72]	skipped	applied	[0, 0.0011, 0.0037, 0.0102, 0.0263, 0.0660, 0.1638, 0.4051, 1.0000]
0 to 62.72, 8 bins	[0, 7.84, 15.68, 23.52, 31.36, 39.20, 47.04, 54.88, 62.72]	applied	applied	[0, 0.1188, 0.2393, 0.3616, 0.4856, 0.6114, 0.7391, 0.8686, 1.0000]
0 to 0.98, 8 bins	[0, 0.1225, 0.245, 0.3675, 0.49, 0.6125, 0.735, 0.8575, 0.98]	applied	applied	[0, 0.1189, 0.2395, 0.3618, 0.4859, 0.6117, 0.7393, 0.8687, 1.0000]
0 to 0.98, 8 bins	[0, 0.1225, 0.245, 0.3675, 0.49, 0.6125, 0.735, 0.8575, 0.98]	applied	skipped	[0, 0.125, 0.25, 0.375 0.5, 0.625, 0.75, 0.875, 1.0]

Table 3.1.2. Bin edge parameters and binning methods. The bin edges we chose are highlighted in blue.

3.2. Sound Distribution

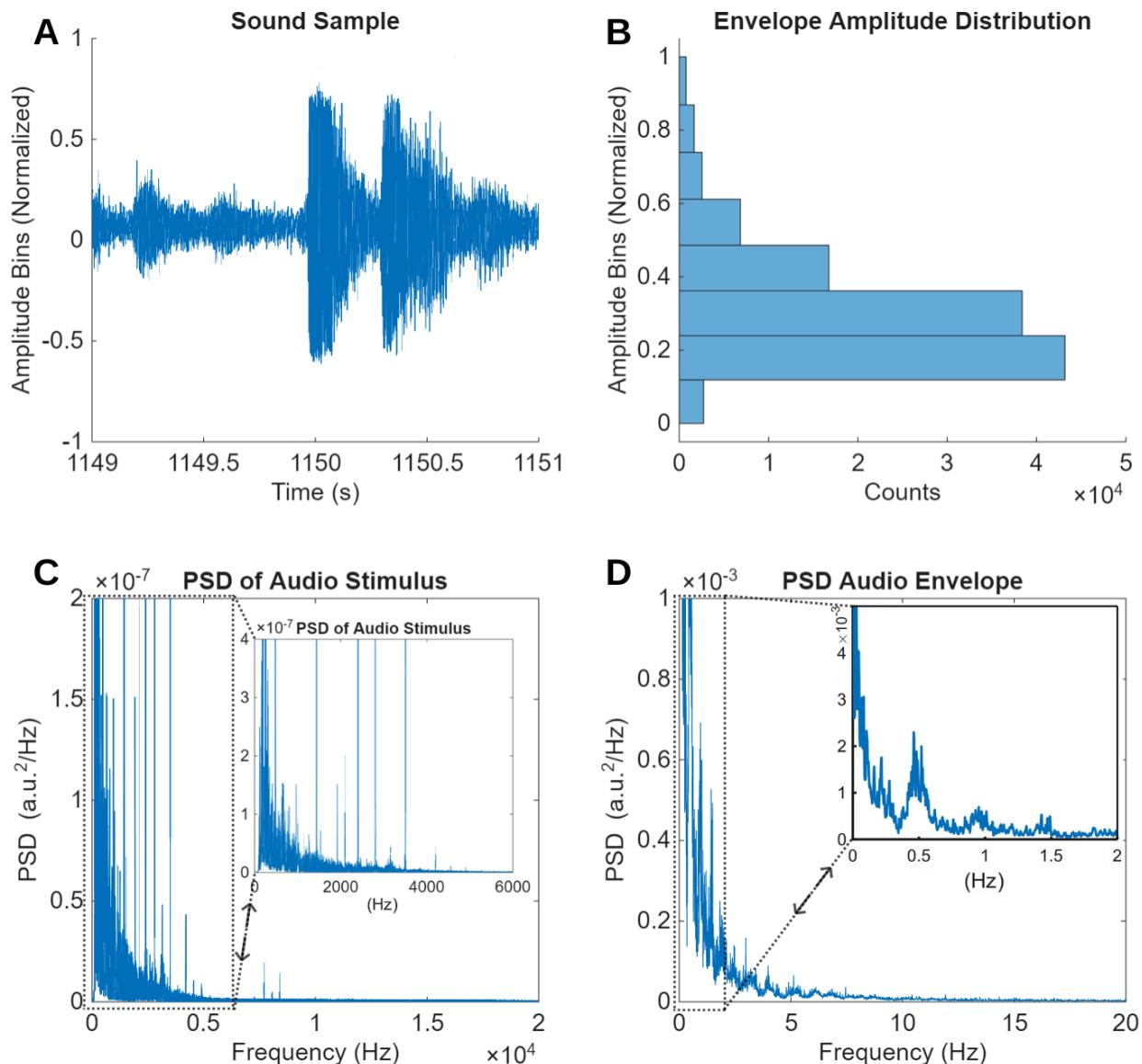
For the derived parameters we investigated the binning of the envelope. For this the averages were calculated across participants for each condition (narrow and wide) and then combined, having a group- and condition-averaged representation that showed prominent group trends.

The audio stimuli used in this study featured a naturalistic soundscape, a mixture of speech and noise. Also, the strong noise factor was observable from audio signal figure and which exhibited an observable strong noise factor (Figure 3.2A). The detailed distribution analysis of sound envelope amplitude shows that the distribution of sound amplitudes in the envelope is skewed towards the lower bins (Figure 3.2B). Specifically, for the 8-bin AB envelope, the amplitude distribution showed the highest density in the 2nd and 3rd bins, corresponding to the amplitude edges of 7.84–15.68 dB and 15.68–23.52 dB, respectively. Interestingly, the lowest bin contained the highest amplitude bins, while the topmost bin included the lowest amplitudes. Overall, this sound amplitude distribution was similar across both narrow and wide conditions. (Appendix, A1).

From the Power Spectral Density (PSD) plots, we observed that the audio signal was largely concentrated at frequencies below 4000 Hz. On top of the main distribution, we detected sharp (instant), outlier peaks around, which might correspond to instant (sharp) alarm and beep sounds, and possibly their harmonics. (Figure 3.2C). These sharp traces can sometimes occur when artificial tones are added, as they may introduce similar sharp frequency components.

PSD of naturalistic sound envelopes, showing a bottom-heavy (right-skewed) frequency distribution, reveals the dominance of low-frequency modulation rates. (Figure 3.2D)

Figure 3.2. Mean Histogram over participants. A. Example audio stimuli segment. B. Amplitude histogram of the entire audio distribution. C. Power spectral densities (PSDs) of the naturalistic sound stimuli. D. PSD of the envelope.



3.3. Channel selection - Topoplot

EEG prediction accuracies vary across channels, due to the stimulus representation relevance at the channels. Initially, the TRF analysis was conducted using all channels, and prediction accuracies were averaged across all participants. For both the narrow and wide conditions, the highest prediction accuracies for the standard and AB envelope models were observed over frontal, frontocentral, and temporal regions. Using the 10-20 Electrode System (Jasper, 1958), the top 5 channels (of 22), - FC2, FC1, C3, C4 and Cz - comprise one midline electrode and four central electrodes from the left and right hemispheres, chosen for their consistently superior performance over all participants. (Figure 3.3)

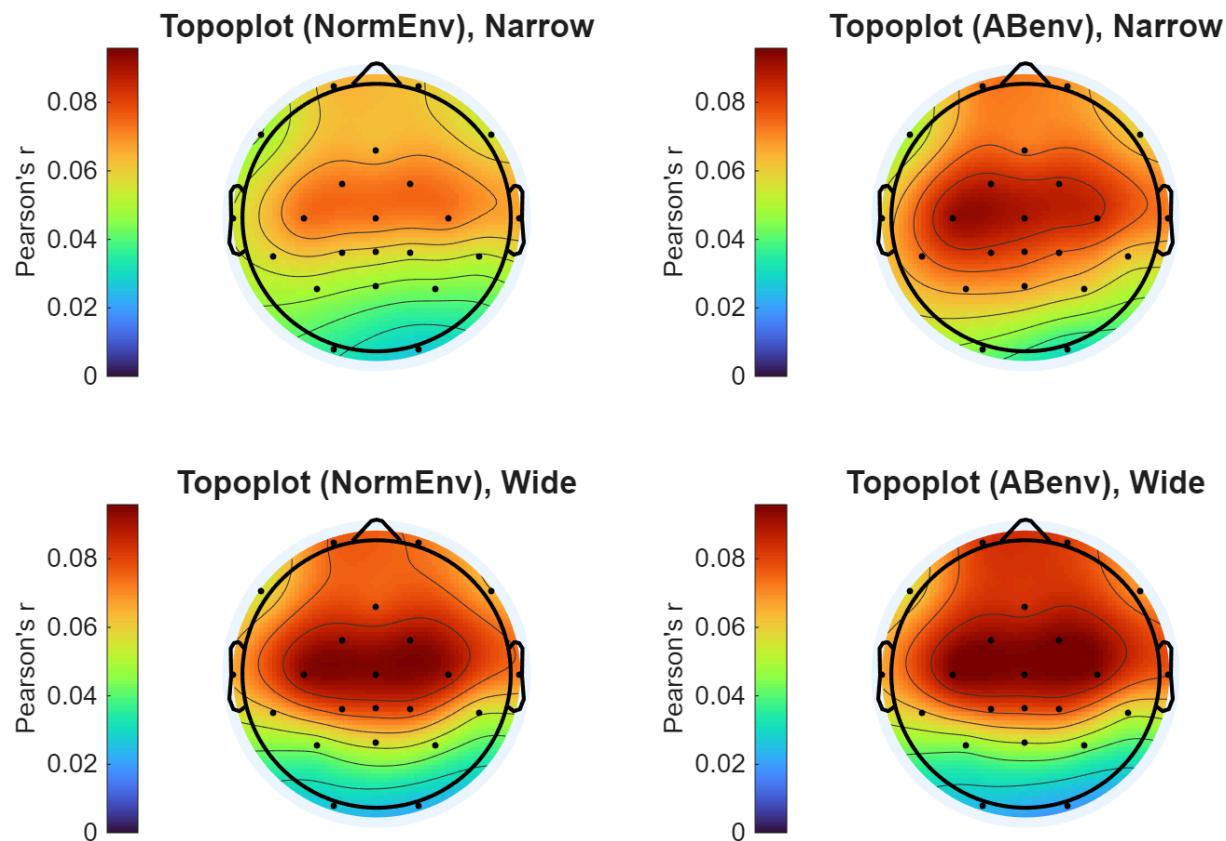


Figure 3.3. Topographic plots of prediction accuracies for the envelope models, averaged across all participants. The top row represents the narrow condition, while the bottom row represents the wide condition. The left plots display results for the standard envelope model, while the right plots show results for the AB envelope model.

3.4. Prediction Accuracies: Condition averaged

For our naturalistic sound dataset, prediction accuracies were calculated for each model and subject (Fig. 3.4). Both stimulus representations (Standard Envelope and AB Envelope) predicted EEG responses with accuracies significantly above the null hypothesis value of 0.05, determined via permutation tests. This demonstrates that the prediction accuracies for both models significantly exceed what would be expected by chance. The AB Envelope model significantly outperformed the Standard Envelope model, with a moderate effect size (paired t-test: $t(20)$, $p = 0.0191$, $d = 0.57252$; Wilcoxon Signed-Rank Test: $p = 0.0187$).

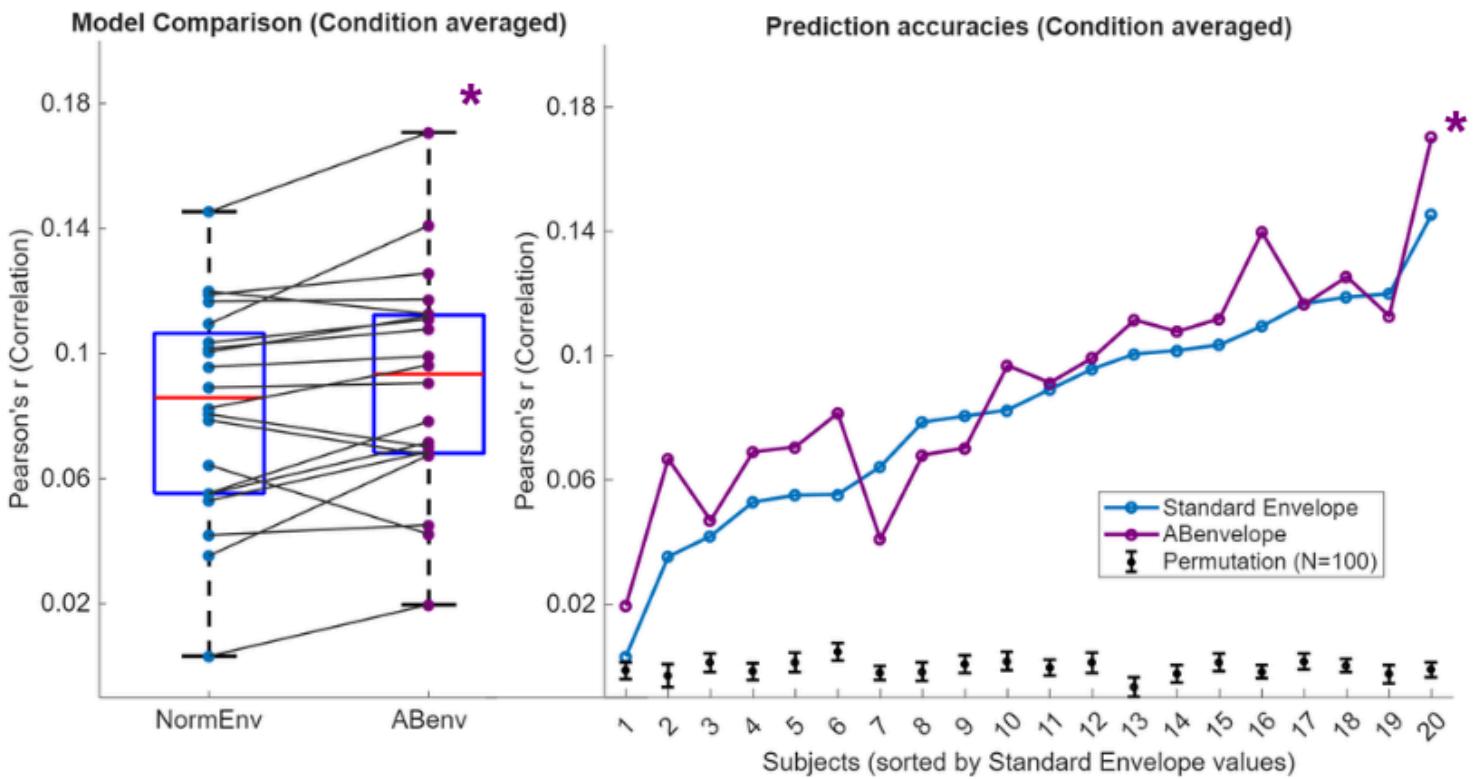


Figure 3.4. Condition averaged prediction accuracies for the envelope models. The right plot presents a box plot summarizing the distribution and mean values for the two models, with each dot representing a participant's prediction accuracy. The left plot displays the individual prediction accuracies for the Standard Envelope (blue) and AB Envelope (purple) models across participants. Black bars indicate the 95% confidence interval derived from the permutation test, which involved 100 repetitions of randomly circular shifted standard envelopes. Asterisk indicates comparisons between AB envelope and Standard envelope, respectively. *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$.

3.5. Weights, Latencies, P2-N1 peaks (Group and Condition Averaged)

To investigate amplitude-dependent changes in TRFs, we analyzed group- and condition-averaged TRF weights, focusing on their magnitude and latency. Specifically, we examined the N1 and P2 peaks, as these peaks are widely recognized markers of sensory and cognitive processing.

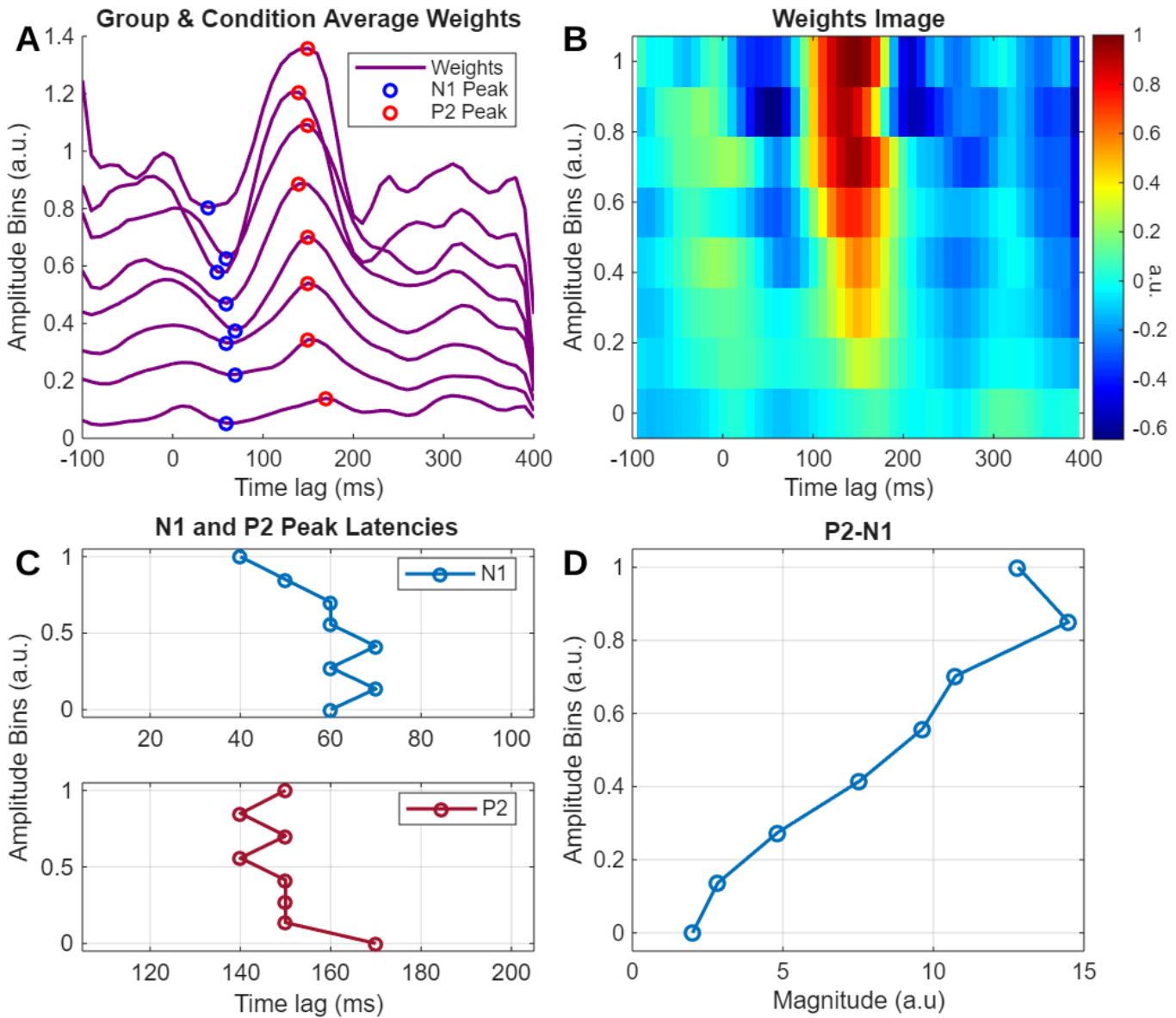
Overall, as stimulus amplitude decreases, the TRF magnitudes (P2-N1) decreases, the latencies of P2 and N1 slightly increases. Additionally, the fluctuations in the higher bin amplitudes become more apparent than those in the lower bins.

To quantify the relationship between stimulus amplitude and N1 latency, the N1 peak in each bin was identified as the largest negative peak in the TRF between 20 and 150 ms. A line was fit to the data, showing an increase of 2.6 ms in N1 latency for each step decrease in amplitude bin ($R^2 = 0.55$, $p = 0.034$). However, the moderate fit of the data suggests this relationship should be interpreted with caution. The P1 peak was not consistently observed, so the focus shifted to the P2 peak, defined as the largest positive peak in the TRF between 70 and 200 ms. For the P2 peak, a line was similarly fit to the data, showing an increase of 2.0 ms in latency per bin step ($R^2 = 0.39$, $p = 0.101$). However, the poor linear fit for P2 also indicates that the relationship is not robust. Overall, no strong or consistent linear relationship between stimulus amplitude and peak latencies was observed.

A similar analysis was conducted to quantify the relationship between stimulus amplitude and TRF magnitude, (with the corresponding P2-N1 peak-to-peak amplitudes shown in (Fig. 3.5D). A strong linear relationship between magnitude and amplitude was observed, with the data fitting well to a line ($R^2 = 0.95$, $p < 0.001$).

This analysis underscores the advantage of the AB envelope-generated TRF model in capturing amplitude-dependent changes, such as latency and magnitude, for each defined amplitude bin—an option not available with standard envelope models. This approach provides more detailed insights into how neural responses vary with stimulus amplitude.

Figure 3.5. Analysis of Amplitude-Dependent Changes Averaged Across Five Representative Channels and Participants. The figure captures how the TRF evolves with varying stimulus amplitudes. The top left panel (A) displays the group-averaged TRF weights for different amplitude bins, while the top right panel (B) provides a visual representation of the same data in an image format. The bottom left panel (C) tracks changes in the latencies of the N1 and P2 peaks. (D) Finally, the bottom right panel (D) quantifies the P2-N1 peak-to-peak magnitude difference, illustrating amplitude-dependent changes in TRF magnitude.



3.6. Index Shuffled Binned-Envelope

To test the model further, after the AB envelope generated, indexes of bins are shuffled. As a result of shuffled bin indices (e.g., swapping the 1st bin with the 5th), the results showed no significant drop in prediction accuracy. In fact, the shuffled bin indices AB envelope and the standard AB envelope exhibited the same performance, suggesting that the model's prediction accuracy was not affected by changes in the bin assignments.

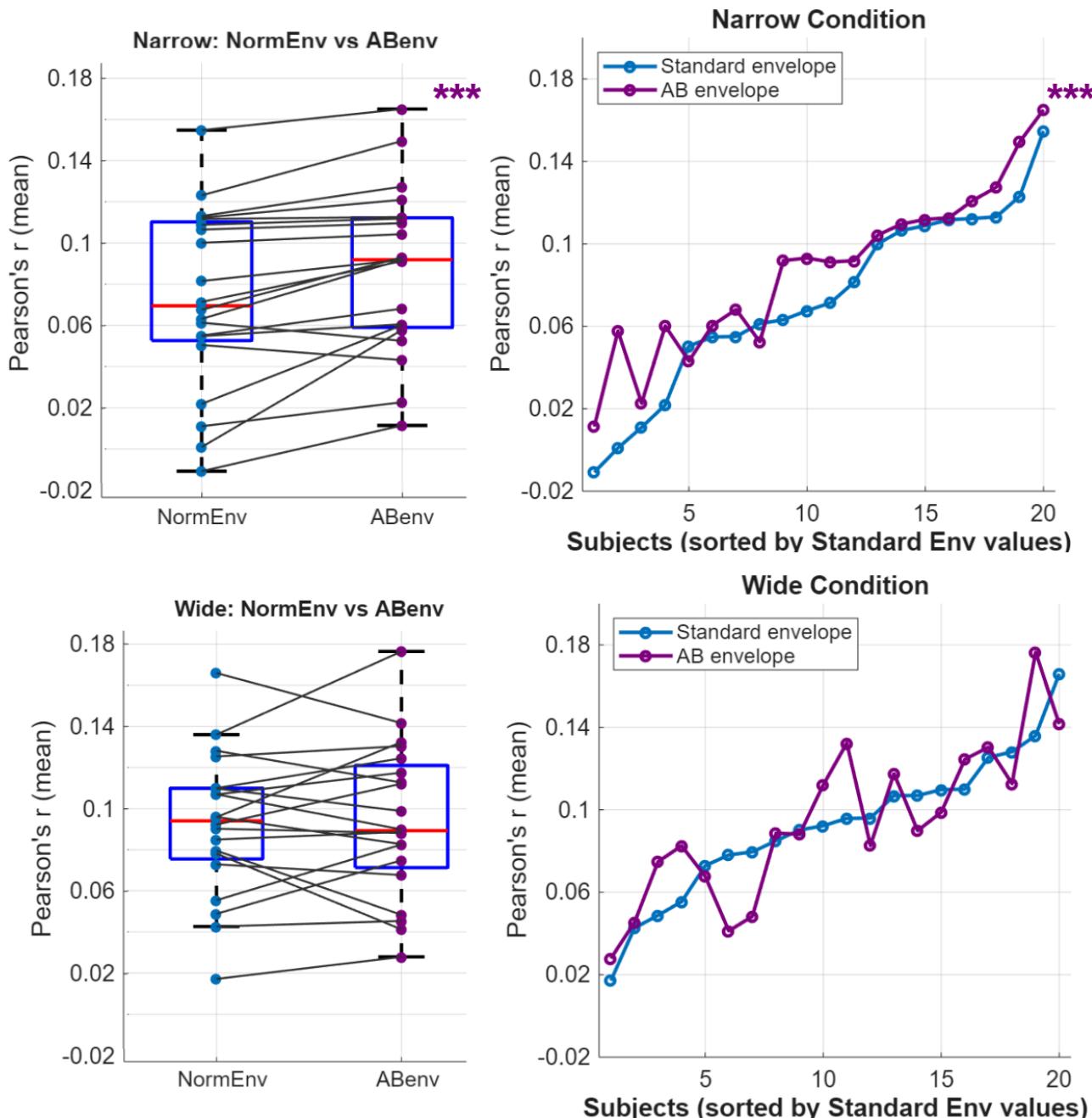
3.7. Impact of Attention Conditions on AB Envelope Modeling

Prediction accuracies were further analyzed separately for the narrow and wide conditions to assess whether attention conditions affect the modeling capabilities of the AB Envelope based modeling. Same as before, a significance criterion of $p < 0.05$ was used to evaluate differences between the Standard Envelope and AB Envelope models.

Prediction accuracies were further analyzed separately for the narrow and wide conditions to assess whether attention conditions affect modeling capabilities or AB envelope modeling. A significance criterion of $p < 0.05$ was used to evaluate differences between the Standard Envelope and AB Envelope models.

In the narrow condition, a significant difference was observed between the Standard Envelope and AB Envelope models (paired t-test: $t = 4.1059$, $p < 0.0006$, $d = 0.91812$; Wilcoxon Signed-Rank Test: $p < 0.0006$). This indicates that the AB Envelope model provided a notably better prediction of EEG responses compared to the Standard Envelope model in this condition. (Figure 3.7A,B.) In contrast, for the wide condition, no significant difference was found between the two models (paired t-test: $t = 0.42$, $p = 0.68$, Cohen's $d = 0.094$; Wilcoxon Signed-Rank Test: $p = 0.74$). This suggests that, under the wide condition, both models performed similarly in predicting EEG responses. (Figure 3.7C,D.)

Figure 3.7. Prediction Accuracies of the Complex Audio Dataset Across Conditions. The figure presents a comparative analysis of response values for the Standard Envelope and AB Envelope models under narrow and wide conditions. The top row depicts the narrow condition, with boxplots for each model shown on the left and individual prediction accuracies plotted on the right. Similarly, the bottom row illustrates the wide condition, following the same format: boxplots on the left and individual prediction accuracies on the right. Top and bottom rows of asterisks indicate comparisons between AB envelope and Standard envelope, respectively. *** $p<0.001$, ** $p<0.01$, * $p<0.05$.



3.8. Combined Model Performance: Condition Averaged

We tested whether combining these two models would be advantageous by combining their stimulus representations and fitting a multivariate TRF (Figure 3.7.). The combination of the AB and Onset Envelope models produced significantly higher prediction accuracies than either the AB or Onset Envelope models alone, with a moderate and large effect size respectively. When averaged across conditions, this combined model showed a statistically significant improvement (vs AB paired t-test(20) = 2.4292 p=0.0252, d=0.5432 Wilcoxon Signed-Rank Test: p=0.0228; vs Onset Envelope t-test(20) = 6.164 p<0.0001, d=1.3783 Wilcoxon Signed-Rank Test: p=0.0002).

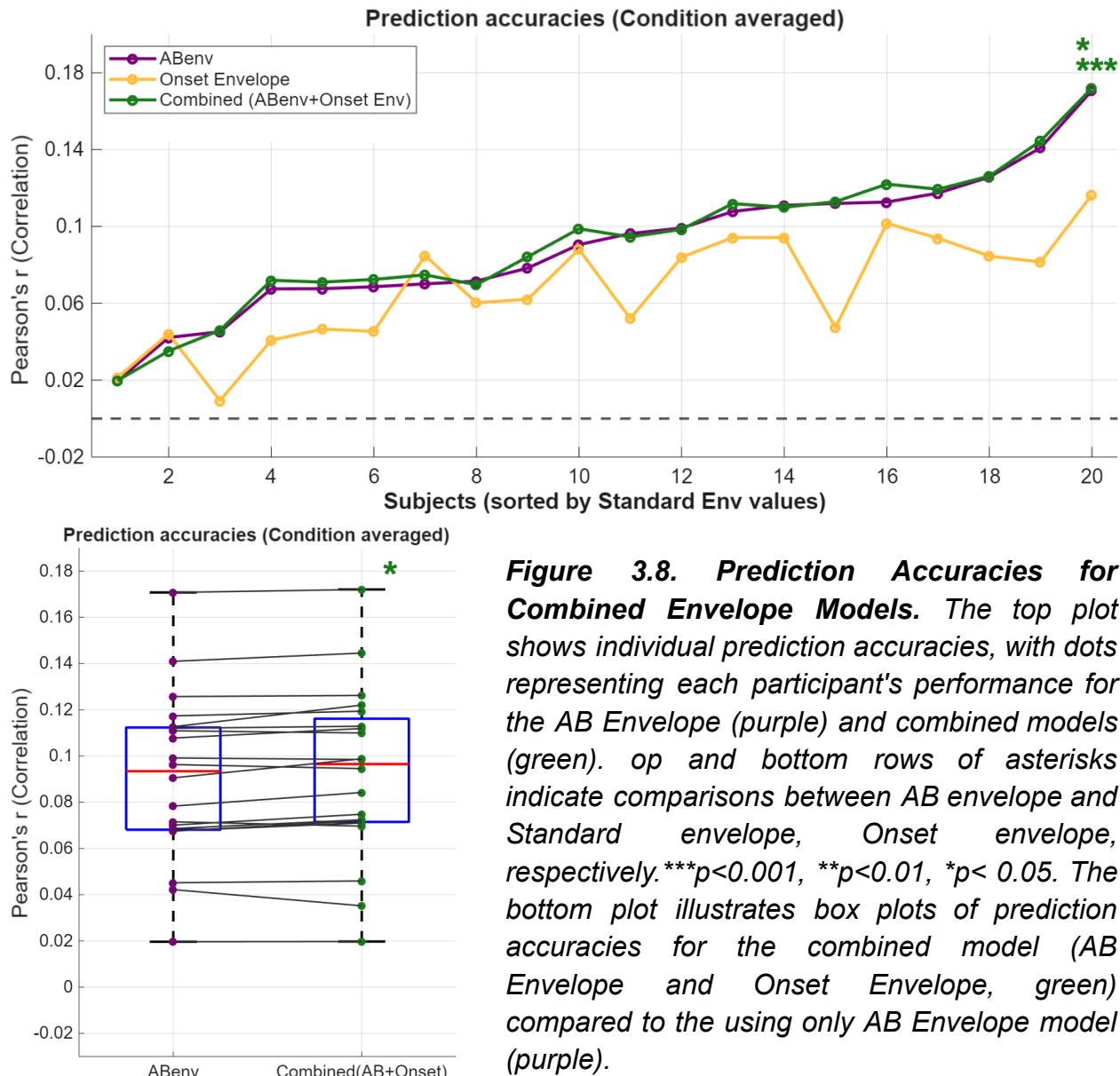


Figure 3.8. Prediction Accuracies for Combined Envelope Models. The top plot shows individual prediction accuracies, with dots representing each participant's performance for the AB Envelope (purple) and combined models (green). Top and bottom rows of asterisks indicate comparisons between AB envelope and Standard envelope, Onset envelope, respectively. ***p<0.001, **p<0.01, *p<0.05. The bottom plot illustrates box plots of prediction accuracies for the combined model (AB Envelope and Onset Envelope, green) compared to the using only AB Envelope model (purple).

4. Discussion

In this study, we explored the potential of amplitude-binned (AB) envelopes as a stimulus representation to model EEG responses to continuous, complex - naturalistic auditory stimuli, aiming to improve prediction accuracy over broadband signal representation methods.

Our results show that binning the envelope into specific amplitude ranges and then using it to fit a multivariate temporal response function (mTRF) for complex - naturalistic audio context, resulted in improved modelling performance compared to the standard envelope. Although the AB envelope is derived from the same data without altering the original information, it adds a structural representation by presenting features in separate amplitude bins. This approach provides additional information by explicitly accounting for amplitude-related variations during the model-building process.

This improvement was observed in the condition-averaged and narrow attention focus conditions, but not in the wide condition. Further combining the AB envelope with onset envelope models led to additional improvements in modelling performance, with a moderate effect size.

Bin Parameters and Binning Methods

As we observed, not only the choice of binning parameters and binning method is critical, but also normalization steps also determine how well the model can capture response dynamics.

Our grid search analysis identified optimal binning parameters for the narrow-attention condition as a lower bound of 0 and upper bound of 0.98, with 8 equal bins, normalized over a range of 0 to 62.72 dB. These parameters differed from Drennan et al. (2019), likely due to the noisier nature of our soundscape stimulus. No optimal parameters were found for the wide-attention condition, highlighting the significant role attention plays in shaping the AB envelope's performance.

Our findings suggest that alternative binning methods, such as linear binning, may be more effective than logarithmic binning, depending on stimulus characteristics and the normalization process. The timing of normalization (before or after the conversion formula, or both) also significantly impacted the model's performance, with pre-conversion normalization producing significant improvement. Overall, these results emphasize the importance of effective normalization and binning strategies in optimizing TRF model performance, especially when dealing with complex stimuli like soundscapes. Exploring these binning strategies is essential for optimizing the predictive accuracy of TRF models.

Tracking Latency and Magnitude Changes

In speech processing, models based on spectrograms, phonemes, and phonetic features have been found to outperform those that use the standard envelope (Di Liberto et al., 2015). However, for both spectrograms and envelopes, the same assumption holds: the TRF morphology remains unchanged. Drennan & Lalor (2019) demonstrated that amplitude-binned envelopes can enhance EEG predictions by better capturing changes in response magnitude and latency associated with stimulus amplitude. The key insight here is that binning the envelope allows for a more detailed examination of the envelope in sections (bin by bin), potentially capturing the brain's non-linear responses to varying stimulus intensities, albeit with a series of linear assumptions.

Drennan et al's (2019) study found that amplitude decreases resulted in changes to the morphology, with a linear increase in latency, but a non-linear relationship with magnitude as amplitude decreased. We observed similar results, where the morphology changed across amplitude bins, as reflected in the varying TRF weights. However, we noticed that as amplitude decreased, magnitude not only decreased significantly but also followed a linear trend, while latency increased. Note. the exact relationship between amplitude reduction and latency changes does not follow a linear pattern, and it remains unclear, requiring further investigation.

Therefore, using the AB envelope enhances the capability of the standard envelope by capturing amplitude-related changes in latency and magnitude in finer amplitude steps, providing a more detailed understanding of how neural responses vary with stimulus amplitude.

On Binned envelope model performance (H1)

Our first hypothesis (H1) proposed that the AB envelope is a better representation than the standard envelope, as it shows improved correlation performance. This hypothesis is mostly supported, as the AB envelope model demonstrates better prediction accuracy compared to the standard envelope model. Specifically, it exhibits a moderate effect size in the condition-averaged version and a high effect size in the narrow condition. However, no significant improvement was observed in the wide-attention condition.

To further validate the robustness of the AB envelope model, we shuffled the bins in our analysis of condition averaged version. This demonstrated that each amplitude bin in the AB envelope model provides meaningful information, and the model's performance remains robust even when the relationship between amplitude levels and neural responses is disrupted. This highlights the advantage of using amplitude bins to capture

nuanced variations in neural responses, enhancing the model's ability to represent complex auditory stimuli.

The results suggest that the performance improvements of the AB envelope over the standard envelope in narrow-focus conditions are likely influenced by attention induced effect, not solely the distribution of audio intensities. To determine which factors might play a role, we investigated both the acoustic profiles of the soundscapes for each condition, as well the effect of attention on processing of sounds.

Explanation via sound distribution: We first looked at sound distribution and the sound intensity distribution revealed more lower-intensity sounds in our dataset. This is due to noise within the complex auditory environment and as a result, our derived envelope appears less smooth compared to speech-only datasets, such as those used in Drennan et al. (2019). However, from the general sound distribution and the power spectral density of both the sound and the envelope, no observable differences were found between the narrow and wide conditions.

Condition-Specific Differences - Attention effect: To understand it more, it may help to look at how attention can change the EEG prediction accuracies. Previous studies have shown that cognitive focus may influence the brain response and consequently recorded EEG responses.

As demonstrated by Rozenkratz et al.,(2023) attended condition leads to stronger EEG responses, as their ERP was higher at an instructed attention condition. Since we also use the same dataset, difference in attention may have modulated the neural response, resulting in a better model fit of binning the envelope in the narrow condition.

The nature of the wide condition likely contributes to the observed results, as beep sounds may be masked by background noise, making them harder for audio processing tools to identify. Envelope models track only audio amplitude changes, and when beeps blend with noise, their amplitude fluctuations may not be significant enough to stand out, resulting in less information being captured. In contrast, in the narrow condition, target alarm sounds are easier to differentiate, enabling the envelope to track these changes more effectively, leading to improved modeling performance.

The advantage of the AB Envelope model over the Standard Envelope model is more pronounced in the narrow condition, likely because alarm sounds are masked by other sounds in the wide condition. While the TRF model primarily uses audio information to predict sound-driven responses, EEG recordings also reflect attention-driven responses, influenced by the listener's ability to focus on relevant stimuli. This suggests that the models may not fully account for attention-driven neural activity. By computing separate models for narrow and wide conditions, we address this limitation, incorporating context-specific variations in attention. This approach acknowledges the

role of attention in shaping neural responses, even if it cannot fully capture structured attention patterns.

O'Sullivan et al. (2015) reconstructed the stimulus from EEG data and assessed the correlation between the EEG-based reconstructed stimulus and the real stimulus. They revealed a significant relationship between EEG measures of attention and performance on high-level attention tasks compared to low-level attention tasks. Similarly, by separating the narrow and wide conditions, we may implicitly capture attention-driven differences, reflecting the listener's ability to focus on relevant stimuli.

Combination of envelopes (H2)

The second hypothesis also confirmed that combining the AB envelope with the onset envelope model further improves prediction accuracy compared to using either only the onset or AB envelope. The AB envelope and onset envelope models reflect different aspects of envelope tracking. Specifically, the onset envelope model likely captures responses to sound onsets and rapid amplitude changes, whereas the AB envelope model tracks continuous amplitude fluctuations. Therefore, the observed improvement may stem from the complementary nature of these representations.

These results further support Drennan and Lalor's (2019) suggestion that integrating the AB envelope with the onset envelope provides complementary information about envelope tracking that neither model alone fully captures. Possibly, the onset of sounds offers an additional source of information that helps explain neural variability, suggesting its relevance to brain processing and the possibility that it is processed separately.

Future studies

While our results are promising, further exploration of binning strategies, combined stimulus representations, and attention-related factors is necessary to optimize TRF modelling.

One potential direction for future work is the development of 3-dimensional models that incorporate both amplitude-dependent and frequency-dependent changes by adding "frequency-binned" envelope features. For example, such models could differentiate between sounds like an 'alarm' and a 'beep' based on their frequency differences, enhancing the model's ability to distinguish these sounds more effectively. We hypothesize that integrating frequency features alongside amplitude in the AB envelope model could improve its performance, providing a more comprehensive representation of the auditory stimulus and capturing additional nuances in neural responses.

Another study could investigate whether separating speech and non-speech categories further enhances the higher predictive accuracy observed in soundscapes, a direction that some of our colleagues have already begun to explore.

5. Conclusion

In summary, we replicated the results of Drennan et al. (2019) and demonstrated that the AB envelope could serve as a promising alternative to the standard envelope.

1. **AB envelope as a sound feature model (H1):** The first hypothesis is mostly supported, as the AB envelope model provides better prediction accuracy compared to the standard envelope. Our findings emphasize the potential of AB envelopes as a more effective feature representation in modelling EEG responses to complex auditory environments, particularly in noisy, real-world scenarios.
2. **Bin Choice:** Parameter selection, especially binning strategies (range, step size, and method, normalization timing), significantly influences the AB envelope's performance. Additionally, linear binning could be considered an important technique for optimizing the TRF model's effectiveness.
3. **Limits of the Method:** While the AB envelope outperforms the standard envelope in narrow attention conditions, it does not capture attention-driven effects effectively in wider attention conditions where sounds are masked by background noise. This highlights the limitations of envelope models and the need for more advanced approaches that integrate cognitive factors and dynamic attention states.
4. **Combination Hypothesis (H2):** The second hypothesis is also supported, as combining the AB envelope with the Onset envelope model further improves prediction accuracy compared to using either model alone. This integration demonstrates the complementary nature of these envelope representations.
5. **Future Studies:** Future research could integrate further cognitive features into AB envelope models. Developing 3D models that combine amplitude and frequency-dependent envelope features could enhance the representation of auditory stimuli, improving sound differentiation (e.g., alarms vs. beeps). Separating speech and non-speech stimuli in soundscapes may also boost the model's ability to predict neural responses. These advances would deepen our understanding of how auditory stimuli and attention influence EEG responses, improving neural encoding models in complex auditory environments.

These advancements could deepen our understanding of how auditory stimuli and attention interact to influence EEG responses, enhancing the effectiveness of neural encoding models in complex auditory environments.

References:

1. Aertsen, A., & Johannesma, P. I. (1981). The spectro-temporal receptive field. A functional characteristic of auditory neurons. *Biological Cybernetics*, 42(2), 133–143.
2. Beagley, H. A., & Knight, J. J. (1967). Changes in auditory evoked response with intensity. *Journal of Laryngology & Otology*, 81(11), 861–873.
3. Boynton, G. M., Engel, S. A., Glover, G. H., & Heeger, D. J. (1996). Linear systems analysis of functional magnetic resonance imaging in human V1. *Journal of Neuroscience*, 16(13), 4207–4221.
4. Carandini, M., Demb, J. B., Mante, V., Tolhurst, D. J., Dan, Y., Olshausen, B. A., Gallant, J. L., & Rust, N. C. (2005). Do we know what the early visual system does? *Journal of Neuroscience*, 25(45), 10577–10597.
5. 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, 604.
6. David, S. V., Mesgarani, N., & Shamma, S. A. (2007). Estimating sparse spectro-temporal receptive fields with natural stimuli. *Neural Computation*, 18(1), 191–212.
7. Delorme, A., & Makeig, S. (2004). EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics. *Journal of Neuroscience Methods*, 13, 9–21. <https://doi.org/10.1016/j.jneumeth.2003.10.009>
8. 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(18), 2457–2465.
9. Drennan, W. R., & Lalor, E. C. (2019). Cortical tracking of complex sound envelopes: Modeling the changes in response with intensity.
10. Giraud, A. L., & Poeppel, D. (2012). Cortical oscillations and speech processing: Emerging computational principles and operations. *Nature Neuroscience*, 15(5), 511–516.
11. Jasper, H. H. (1958). The Ten-Twenty Electrode System of the International Federation. *Electroencephalography and Clinical Neurophysiology*, 10(4), 371–375.
12. Kalarus, M. (2021). Tetris 3D. Available online at: <https://www.mathworks.com/matlabcentral/fileexchange/92243-tetris3d>
13. Kayser, H., Ewert, S. D., Anemüller, J., Rohdenburg, T., Hohmann, V., & Kollmeier, B. (2009). Database of multichannel in-ear and behind-the-ear head related and binaural room impulse responses. *EURASIP Journal on Advances in Signal Processing*, 2009, 298605. <https://doi.org/10.1155/2009/298605>

14. Lalor, E. C. (2019). Evoked activity plays a very substantial role in the cortical tracking of natural speech. Annual Meeting of the Cognitive Neuroscience Society, March 23–26, San Francisco, CA. Abstract C111.
15. 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.
16. Liberty, S. H., & Huth, A. G. (2020). The revolution will not be controlled: Natural stimuli in speech neuroscience. *Journal of Cognitive Neuroscience*, 32(10), 1896–1908. <https://doi.org/10.1080/23273798.2018.1499946>
17. Luo, H., & Poeppel, D. (2007). Phase patterns of neuronal responses reliably discriminate speech in human auditory cortex. *Neuron*, 54(6), 1001–1010. <https://doi.org/10.1016/j.neuron.2007.06.004>
18. Machens, C. K., Wehr, M. S., & Zador, A. M. (2004). Linearity of cortical receptive fields measured with natural sounds. *Journal of Neuroscience*, 24(5), 1089–1100.
19. Mesgarani, N., David, S. V., Fritz, J. B., & Shamma, S. A. (2009). Influence of context and behavior on stimulus reconstruction from neural activity in primary auditory cortex. *Journal of Neurophysiology*, 102(6), 3329–3339.
20. O'Sullivan, J. A., Power, A. J., Mesgarani, N., Rajaram, S., Foxe, J. J., Shinn-Cunningham, B. G., Slaney, M., Shamma, S. A., & Lalor, E. C. (2015). Attentional selection in a cocktail party environment can be decoded from single-trial EEG. *Cerebral Cortex*, 25(5), 1697–1706.
21. Perrault, N., & Picton, T. W. (1984). Event-related potentials recorded from the scalp and nasopharynx. I. N1 and P2. *Electroencephalography and Clinical Neurophysiology/Evoked Potentials Section*, 59(3), 177–194. [https://doi.org/10.1016/0168-5597\(84\)90058-3](https://doi.org/10.1016/0168-5597(84)90058-3)
22. Rosenkranz, M., Haupt, T., & Bleichner, M. G. (2023). Investigating the attentional focus to workplace-related soundscapes in a complex audiovisual-motor task using EEG. *Neurophysiological Perspectives*.
23. Theunissen, F. E., & Elie, J. E. (2014). Neural processing of natural sounds. *Nature Reviews Neuroscience*, 15(9), 506–516.
24. Theunissen, F. E., David, S. V., Singh, N. C., Hsu, A., Vinje, W. E., Gallant, J. L. (2001). Estimating spatio-temporal receptive fields of auditory and visual neurons from their responses to natural stimuli. *Network: Computation in Neural Systems*, 12(3), 289–316.
25. Wu, M. K., David, S. V., & Gallant, J. L. (2006). Complete functional characterization of sensory neurons by system identification. *Annual Review of Neuroscience*, 29, 477–505.
26. Zhang, Y., & Li, W. (2016). Exploring relevant features for EEG-based investigation of sound perception in naturalistic soundscapes. *OSF Preprints*. <https://doi.org/10.31234/osf.io>