



## OPEN ACCESS

EDITED BY  
Edmund C. Lalor,  
University of Rochester, United States

REVIEWED BY  
Xiangbin Teng,  
Max Planck Institute for Empirical  
Aesthetics, MPG, Germany  
Nathaniel J. Zuk,  
University of Rochester, United States

\*CORRESPONDENCE  
Wouter David  
wouter.david@kuleuven.be

SPECIALTY SECTION  
This article was submitted to  
Neuro-Otology,  
a section of the journal  
Frontiers in Neurology

RECEIVED 10 January 2022  
ACCEPTED 29 June 2022  
PUBLISHED 03 August 2022

CITATION  
David W, Gransier R and Wouters J  
(2022) Evaluation of phase-locking to  
parameterized speech envelopes.  
*Front. Neurol.* 13:852030.  
doi: 10.3389/fneur.2022.852030

COPYRIGHT  
© 2022 David, Gransier and Wouters.  
This is an open-access article  
distributed under the terms of the  
[Creative Commons Attribution License  
\(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or  
reproduction in other forums is  
permitted, provided the original  
author(s) and the copyright owner(s)  
are credited and that the original  
publication in this journal is cited, in  
accordance with accepted academic  
practice. No use, distribution or  
reproduction is permitted which does  
not comply with these terms.

# Evaluation of phase-locking to parameterized speech envelopes

Wouter David\*, Robin Gransier and Jan Wouters

ExpORL, Department of Neurosciences, KU Leuven, Leuven, Belgium

Humans rely on the temporal processing ability of the auditory system to perceive speech during everyday communication. The temporal envelope of speech is essential for speech perception, particularly envelope modulations below 20 Hz. In the literature, the neural representation of this speech envelope is usually investigated by recording neural phase-locked responses to speech stimuli. However, these phase-locked responses are not only associated with envelope modulation processing, but also with processing of linguistic information at a higher-order level when speech is comprehended. It is thus difficult to disentangle the responses into components from the acoustic envelope itself and the linguistic structures in speech (such as words, phrases and sentences). Another way to investigate neural modulation processing is to use sinusoidal amplitude-modulated stimuli at different modulation frequencies to obtain the temporal modulation transfer function. However, these transfer functions are considerably variable across modulation frequencies and individual listeners. To tackle the issues of both speech and sinusoidal amplitude-modulated stimuli, the recently introduced Temporal Speech Envelope Tracking (TEMPEST) framework proposed the use of stimuli with a distribution of envelope modulations. The framework aims to assess the brain's capability to process temporal envelopes in different frequency bands using stimuli with speech-like envelope modulations. In this study, we provide a proof-of-concept of the framework using stimuli with modulation frequency bands around the syllable and phoneme rate in natural speech. We evaluated whether the evoked phase-locked neural activity correlates with the speech-weighted modulation transfer function measured using sinusoidal amplitude-modulated stimuli in normal-hearing listeners. Since many studies on modulation processing employ different metrics and comparing their results is difficult, we included different power- and phase-based metrics and investigate how these metrics relate to each other. Results reveal a strong correspondence across listeners between the neural activity evoked by the speech-like stimuli and the activity evoked by the sinusoidal amplitude-modulated stimuli. Furthermore, strong correspondence was also apparent between each metric, facilitating comparisons between studies using different metrics. These findings indicate the potential of the TEMPEST framework to efficiently assess the neural capability to process temporal envelope modulations within a frequency band that is important for speech perception.

## KEYWORDS

temporal processing, envelope modulations, envelope encoding, auditory steady-state responses (ASSR), speech processing

## Introduction

Natural speech is a complex and dynamic signal. One prominent component of the speech signal is the temporal envelope. The speech envelope contains slow modulations that are related to linguistic information at different timescales such as phrases, words, syllables, and phonemes (1, 2). The modulation spectrum of the speech envelope exhibits a prominent peak for slow modulations of 4–5 Hz (3, 4), which corresponds to the syllable rate in speech (1, 5–7). Since the timescales of these slow modulations coincide with spoken syllables, access to these envelope modulations and their representation in the neural signal traveling through the auditory pathway is essential for speech perception, especially when access to spectral information is limited (8–12).

Two main electrophysiological paradigms are often used to investigate the neural representation of these slow envelope modulations throughout the auditory pathway. One paradigm involves neural entrainment to speech, which refers to cortical responses that consistently phase-lock to slow modulations of the speech envelope (13). The relation between neural responses and the speech envelope through phase-locking has been established with magneto- and electroencephalography (MEG/EEG) (14–16). While listening to speech, the phase pattern of the neural response is consistent with the speech envelope modulations of 4–8 Hz (17, 18). Interestingly, several studies suggested that speech perception performance is associated with the degree of phase-locking to the speech envelope (19–21). In other words, neural phase-locked patterns that are less consistent with the speech envelope are associated with degraded speech perception. For example, higher disruption of neural phase-locking during listening with electrical transcranial stimulation has been shown to result in more degraded speech perception (22). These findings suggest that phase-locking to the speech envelope in the auditory pathway plays an important role in speech perception. Moreover, hierarchical linguistic structures – such as words, phrases, and sentences – are differentiated by input acoustical cues and linguistic higher-order comprehension processes (23–25). The phase-locked responses to speech from the auditory pathway consist of cortical activity at different timescales (or modulation frequency bands) that concurrently track different linguistic structures at different hierarchical levels.

Analyses of phase-locked responses to speech have pointed to distinct functional roles of the delta (1–4 Hz) and theta (4–8 Hz) bands. On the one hand, phase-locking in the delta band is largely associated with the amount of linguistic information in the speech signal (26, 27) and with the listener's proficiency in the language (28–30). By manipulating the different levels of linguistic structure in the speech signals, this can be studied. When listening to a stream of synthesized Chinese sentences, in which the sentence rate was not present in the envelope but was encoded in the linguistic structure, native Chinese listeners

did show phase-locking at the sentence rate while native English listeners did not (29). Neural phase-locking is also associated with lexical, syntactic, and/or semantic changes in the linguistic content when the speech is comprehended. The theta band (4–8 Hz), on the other hand, seems to be more dependent on the saliency of the perceived acoustic envelope. To assess how envelope modulations at these low frequencies are processed by the auditory system, one can use techniques that alter the linguistic content of speech. Distortions to the speech signal can consequently also affect the linguistic message conveyed (31, 32). These findings show that the envelope and the linguistic content of speech are interdependent (13, 33–35). However, the relative contributions to neural phase-locked responses of the speech envelope on the one hand and the linguistic content of speech, on the other hand, are difficult to disentangle from each other. Several studies have shown the applicability to use amplitude-modulated (AM) stimuli to assess phase-locked responses to envelope modulations (36–39).

Sinusoidally amplitude-modulated (SAM) stimuli are at the basis of the other paradigm to investigate the neural representation of envelope modulations. These stimuli evoke auditory steady-state responses (ASSR) (40) of which the strength reflects the ability of the auditory pathway to phase-lock to the stimulus' modulation frequency (i.e., the response is synchronized to the envelope fluctuations). ASSRs evoked by stimuli with modulations below 20 Hz originate predominately from the auditory cortex, while those evoked with higher frequencies originate from subcortical and brainstem regions (41–44). Studies have indicated that speech perception performance in noise is correlated with 40-Hz ASSRs (45–47) and 80-Hz ASSRs (47–49). In addition, ASSRs elicited by 20-Hz and 4-Hz modulations are associated with phoneme and sentence scores, respectively (48–50). To obtain a sense of the overall capacity of neural modulation processing, ASSRs are measured over a wide range of modulation frequencies. The ASSR amplitude as a function of modulation frequency is the temporal modulation transfer function (TMTF). The TMTF shows a broad peak around 80 and 40 Hz (36–39), and also around 20 Hz (36). Interestingly, the TMTF shows large variations in ASSR evoked by modulation frequencies below 20 Hz and across listeners (36). Therefore, to gain insight into the overall processing capacity of these slow modulations, one would have to measure several ASSRs within this range to evaluate the overall capability to process speech-relevant modulations. However, this approach is time-consuming and could potentially be performed more efficiently using a speech-like stimulus that contains the modulation frequencies of interest.

To overcome the issues that are encountered with speech and SAM stimuli, Gransier and Wouters (51) developed the Temporal Envelope Speech Tracking (TEMPEST) framework. The TEMPEST framework enables the creation of stimuli with parameterized envelopes which can be used to assess the effect

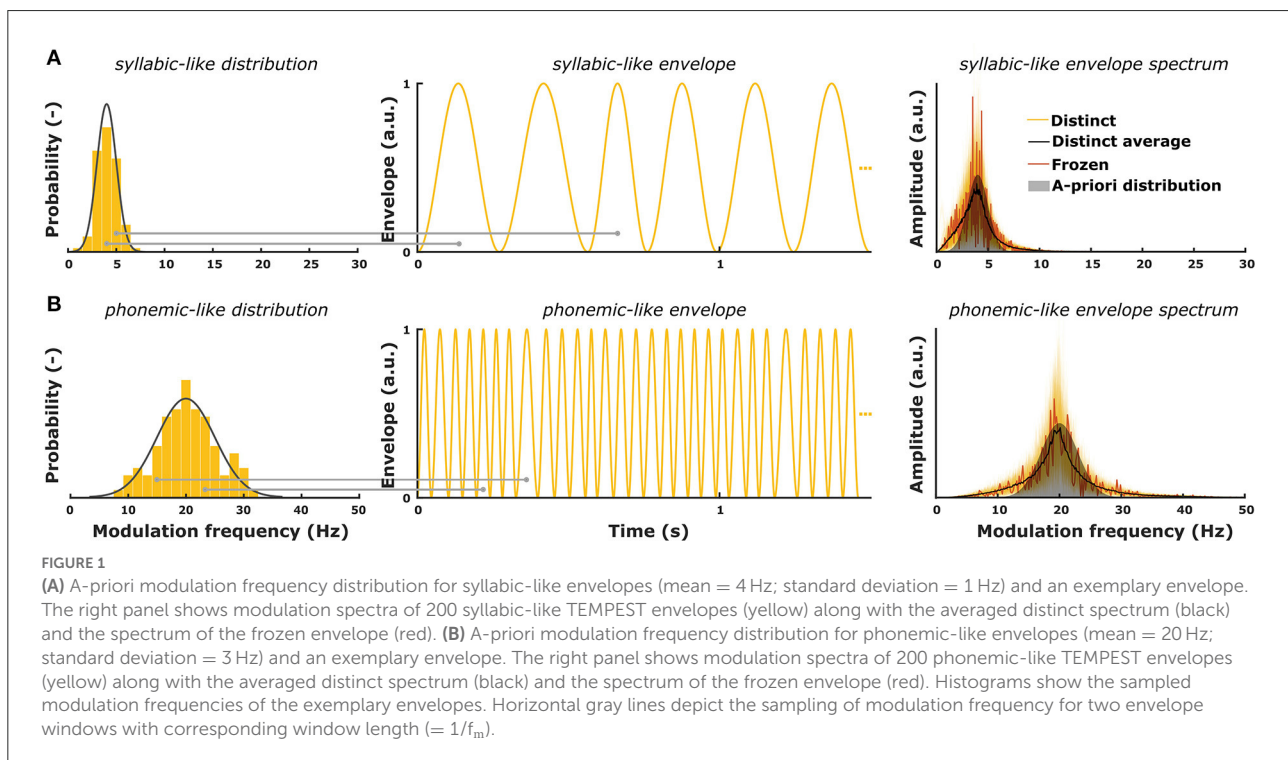
of specific characteristics of the speech envelope on neural processing (e.g., envelopes that contain the same modulations as natural speech). In the present study, we investigate whether TEMPEST-based stimuli that consist of syllabic-like and phonemic-like modulations—as present in natural speech—can be used to gain insight into the speech-weighted electrophysiological TMTF of normal-hearing listeners. To this end, we elicited responses with TEMPEST stimuli based on distributions of modulation frequencies close the syllable ( $\sim 4$  Hz) and phoneme ( $\sim 20$  Hz) rates in speech. Furthermore, we also recorded ASSRs, which are normally used to assess the electrophysiological TMTF, with modulation frequencies that covered the same range as those in the TEMPEST stimuli. We compared the overall activity of the TEMPEST neural responses and that of the ASSRs. We expect that the overall TEMPEST neural activity corresponds to the speech-weighted overall activity within the ASSR TMTF and that the TEMPEST framework can be used to efficiently probe the speech-weighted electrophysiological TMTF in normal-hearing listeners. To this end, we used different power- and phase-based electrophysiological metrics that are widely used in the literature. Many studies make use of various electrophysiological metrics (or terminologies) to characterize phase-locked responses to AM stimuli. Some of the studies made use of power-based metrics [e.g., in Gransier et al. (36), Purcell et al. (37), Poulsen et al. (38)] while other studies applied phase-based metrics [e.g., in Luo and Poeppel (17), Howard and Poeppel (18)]. Due to the use of different metrics, comparing results across studies is difficult. Therefore, we included different

power- and phase-based metrics and investigated how they relate to each other in order to facilitate these comparisons across studies.

## Materials and methods

### TEMPEST framework

Gransier and Wouters (51) introduced the TEMPEST framework in which amplitude-modulated stimuli are created based on an a-priori distribution of modulation frequencies that are relevant for speech. The purpose of the TEMPEST framework is to evaluate the overall envelope encoding ability of the auditory system with stimuli containing a range of envelope modulations. TEMPEST stimuli have a quasi-regular envelope which is generated by concatenating windows over time (Figure 1). Each window in the envelope can represent the occurrence of an acoustic unit in natural speech. The duration of each window depends on random sampling from a probability distribution of modulation frequencies. Each randomly sampled modulation frequency ( $f_m$ ) is inverted to determine the duration ( $T_{\text{window}} = 1/f_m$ ) of subsequent windows (Figure 1). Furthermore, each window can have some fixed or variable parameters, such as peak amplitude, onset time, etc. A simple example is the SAM stimulus, which can be created within the TEMPEST framework using sinusoidal windows with a fixed peak amplitude and only one modulation frequency. The next examples are two TEMPEST stimuli used in this



study (Figure 1, right). These stimuli have different modulation frequency distributions: one centered around 4 Hz (syllable rate) and one around 20 Hz (phoneme rate) (Figure 1, left). Due to sampling of the distributions, the envelope modulation spectrum will also contain these modulation frequencies with a peak at the center frequency. After its generation, the envelope is used to modulate a carrier signal to finalize the creation of the TEMPEST stimulus.

The main goal of this study is to validate whether the TEMPEST framework can be used to assess the speech-weighted electrophysiological TMTF in normal-hearing listeners. The TEMPEST framework would be a useful tool to investigate the overall neural capability to process envelope modulations which can potentially be related to speech perception performance. To this end, we generated “basic” TEMPEST stimuli using a Gaussian probability function of low modulation frequencies that are apparent in the speech envelope.

## Participants

Ten normal-hearing native-Dutch young adults (ages from 19 to 27 years; 3 males and 7 females) participated in this study. No participants had neurological deficits. All participants had normal hearing (pure tone thresholds  $\leq 25$  dB HL for all octave frequencies between 250 and 8,000 Hz). This study was approved by the Medical Ethical Committee of the UZ Leuven hospital (study number: B322201524931). All participants gave written informed consent before participation.

## Stimuli

### SAM stimuli

ASSRs with different modulation frequencies were recorded to obtain individual electrophysiological TMTFs within the modulation frequency ranges of the TEMPEST stimuli. Modulation frequencies of the SAM stimuli were chosen to sample the modulation bands of the TEMPEST stimuli (Figure 1, left). Syllabic-like SAM stimuli with modulation frequencies of 2–6 Hz and phonemic-like SAM stimuli with modulation frequencies of 17–23 Hz were included. All SAM stimuli were created in a custom stimulation software (52). Modulation frequencies were adjusted such that there is an integer number of cycles within one trial of 1.024 s. However, we will further report using rounded modulation frequencies for readability. Modulation depth was set at a maximum of 100% in order to elicit as large ASSRs as possible. The carrier was speech-weighted noise which was generated from the long-term average spectrum of 730 Dutch sentences of the LIST corpus (53). Blocks of 2.56 min were recorded in each measurement session so that 300 trials in total were recorded for each modulation frequency.

### TEMPEST stimuli

TEMPEST envelopes for this study were generated in Matlab R2016b using Hann windows. Hann windows were used because they have a start- and endpoint at zero to prevent discontinuities in the envelope. The peak amplitude of the windows was always at a maximum of 1 such that the effective modulation depth of the TEMPEST stimuli was 100%. We generated two types of TEMPEST stimuli: syllabic-like and phonemic-like stimuli (Figure 1). Modulation distributions of the TEMPEST stimuli were based on modulation rates that are particularly important for speech, i.e., the natural rates of syllables and phonemes (2, 7). The modulation distribution of syllabic-like TEMPEST envelopes closely matched the low envelope modulation spectrum of speech, which shows a peak around 4 Hz (3, 4). The phonemic-like modulation distribution was based on phoneme length statistics in speech from which the mean duration was found to be around 50 ms (54), which corresponds to a center modulation frequency of 20 Hz. The standard deviations of the distributions were 1 Hz and 3 Hz the envelopes of the syllabic-like and phonemic-like TEMPEST stimuli, respectively (Figure 1).

The duration of the syllabic-like and phonemic-like stimuli were 5.12 s and 25.6 s long in order to reach a similar number of envelope windows and to sufficiently sample the modulation distributions. The envelopes were tested for sufficient statistical similarity to the modulation distribution using the Kolmogorov-Smirnov test with a significance level of  $\alpha = 0.05$ . Additionally, we applied criteria to ensure that the envelope modulation sample mean and standard deviation did not deviate too far from those of the a-priori distribution. We used  $\Delta\mu \leq 0.05$  Hz and  $\Delta\sigma \leq 0.05$  Hz for syllabic-like envelopes, and  $\Delta\mu \leq 0.25$  Hz and  $\Delta\sigma \leq 0.1$  Hz for phonemic-like envelopes, with  $\Delta\mu$  the difference between the means and  $\Delta\sigma$  the difference between standard deviations of the sample and a-priori distributions. Envelopes that did not meet these criteria were discarded and new ones were generated instead until they met the criteria. This procedure was continued until 200 syllabic-like and phonemic-like TEMPEST envelopes were obtained. Only 20% of the total amount of generated envelopes passed the test and both criteria. Finally, these envelopes were used to modulate segments of speech-weighted noise based on Dutch LIST sentences (53).

In the main experiment, one single syllabic-like and one phonemic-like stimulus were presented repeatedly to the listener. These stimuli are referred to as frozen stimuli since the same temporal pattern was used over again. The goal of the frozen stimuli was to test robust neural phase-locking and evoked power in the modulation distribution frequency range and to compare this neural activity with ASSRs. Additionally, the remaining syllabic-like and phonemic-like stimuli were presented only once to the listener. Since these stimuli were temporally different from each other, they are referred to as distinct stimuli. Distinct stimuli were used as a baseline measurement with respect to the frozen stimuli (17, 55–57).



The number of distinct stimuli equaled the number of repeated presentations of the frozen stimulus so that bias by differences in the number of trials is minimized (58). Stimuli were presented in blocks of 5.12 min, in which either frozen stimuli were repeated or distinct stimuli were presented in random order. In total, there were 156 frozen and distinct syllabic-like trials (12 presentations in 13 blocks), and 180 frozen and distinct phonemic-like trials (60 presentations in 3 blocks). Each block was preceded with a short 2.56-s TEMPEST segment generated with the same parameters. The evoked neural activity to this segment contains an onset response that would interfere with the main analysis. Therefore, the EEG recordings corresponding to this segment were immediately discarded.

## Equipment

### Calibration and presentation setup

Presentation of all stimuli was done using custom-built software interfacing with an RME-Hammerfall DSP Multiface II soundcard and delivered monaurally through an Etymotic ER-3A insert earphone to the right ear. All stimuli were calibrated using a 2-cc coupler of an artificial ear (Brüel & Kjær, type 4,152) and presented at 70 dB sound pressure level (SPL) at a sampling rate of 32 kHz. Two measurement sessions were conducted whereby each session started with a set of ASSR stimuli in a pseudo-random order which was followed by a set of phonemic-like and syllabic-like TEMPEST stimuli in a pseudo-random order as well.

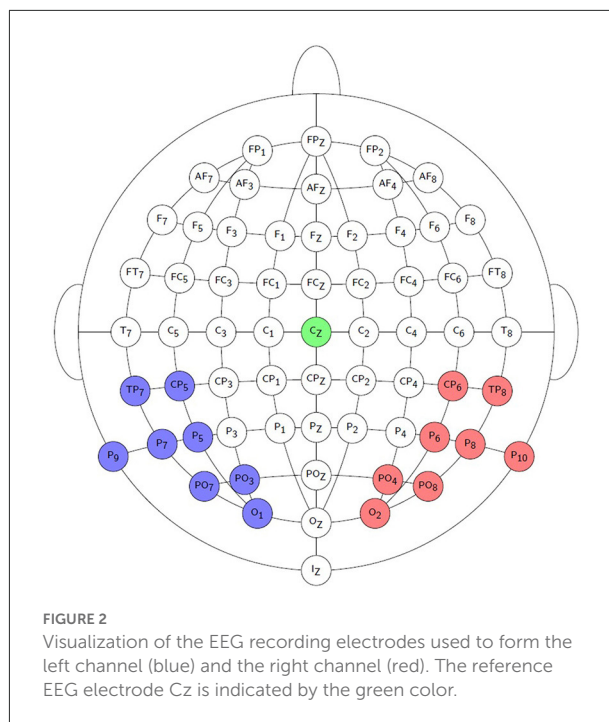
### EEG recording setup

EEG was recorded using a 64-channel BioSemi ActiveTwo recording system with a sampling rate of 8,192 Hz and a recording bandwidth of 0 to 1,683 Hz. A head cap with 64 Ag/AgCl recording electrodes was placed on the scalp of every participant. The electrode positions were placed across the scalp according to the international standard 10–20 system (59). All recordings were made in a double-walled soundproof booth that is equipped with a Faraday cage to avoid signal interference as much as possible. Participants watched a silent movie by choice while seated in a relaxing chair. They were offered a head pillow and asked to move as little as possible to minimize head movement/muscle artifacts.

## Signal processing and response quantification

### Preprocessing

Offline signal processing was done in Matlab R2016b. EEG recordings were high-pass filtered using a 1st order Butterworth filter with a cut-off frequency of 0.5 Hz to remove any DC



**FIGURE 2**  
Visualization of the EEG recording electrodes used to form the left channel (blue) and the right channel (red). The reference EEG electrode Cz is indicated by the green color.

component and slow drifts. Recordings were referenced to electrode Cz by subtracting the recording of Cz from those of the other channels. 5% of the trials were discarded from the analysis based on the highest peak-to-peak amplitudes, as they were assumed to contain muscle and other recording artifacts. Due to measurement errors, not all trials could be obtained from each participant. Only 108 frozen phonemic-like trials could be retained from Participant 7, while 162 phonemic-like trials could be retained in all other cases. The minimum number of retained syllabic-like trials is 115 and the maximum number is 136 across all participants. Time signals of the parieto-occipital recording electrodes were averaged into a left and a right hemispheric channel. Recording electrodes O1, PO3, PO7, P9, P7, P5, CP5, and TP7 formed the left hemispheric channel, while recording electrodes O2, PO4, PO8, P10, P8, P6, CP6, and TP8 formed the right hemispheric channel. See Figure 2 for a visualization of the selected electrodes.

In the case of ASSRs, all 300 trials of each modulation frequency were successfully recorded. Syllabic ASSR ( $f_{\text{mod}} = 2\text{--}6\text{ Hz}$ ) recordings were grouped into sweeps of 5 trials, while phonemic-like ASSR ( $f_{\text{mod}} = 17\text{--}23\text{ Hz}$ ) recordings were grouped into sweeps of 1 trial. Syllabic-like and phonemic-like ASSR sweep lengths were thus 5.12 s and 1.024 s, respectively. Consequently, the number of cycles in each sweep is similar for both syllabic-like and phonemic-like ASSRs in order to have similar phase estimation during analysis. The rest of the preprocessing procedure is the same as for the TEMPEST recordings.

## Neural response analyses

Amplitude and phase for each modulation frequency were extracted from the individual or averaged response trials after transforming into the spectral domain. ASSR sweeps were transformed using the discrete Fourier transform. TEMPEST response trials were transformed into Fourier spectrograms with Hanning windows in which the window length and window overlap were tuned such that phase estimation is similar to that for ASSRs. The window length was equal to the length of the corresponding syllabic-like or phonemic-like ASSR sweep. The window overlap corresponded to three times the reciprocal of the mean modulation frequency in each TEMPEST stimulus such that subsequent windows are, on average, one cycle from each other. Thus, for syllabic-like TEMPEST, spectrograms were computed with 5.12 s window length and 0.25 s window step, whereas for phonemic-like stimuli, a window length of 1.024 s and a window step of 0.05 s were used. Since different spectrogram parameter values were used, the frequency resolution differed between syllabic-like and phonemic-like stimuli. Response bins are 0.195 Hz/bin and 0.977 Hz/bin, respectively. Amplitude and phase were extracted from each time-frequency bin in the spectrogram. These values were used to compute several electrophysiological metrics listed below.

To gain insight into the characteristics and robustness of the recorded neural responses and to compare the TEMPEST responses with ASSRs, four electrophysiological metrics were employed in our analysis. A small selection of metrics have been employed because many different metrics are being used in the literature and this makes comparisons and conclusions across studies more difficult. In order to investigate how different metrics relate to each other and to facilitate comparisons between studies, the metrics used in our analyses represent some of the most widely used ones in the power and phase domain. Two of them are power-based metrics, namely power and signal-to-noise ratio (SNR) of the averaged response. Power is computed after obtaining the amplitude spectrum of the averaged neural response and squaring the amplitude in each frequency bin. This metric reflects the overall neural activity evoked by the stimulus (60). The SNR is taken as the power of the averaged neural response divided by power of the neural background noise. Power of the averaged neural response is computed as the mean power across stimulus trials in each frequency bin. Power of the neural background noise is computed as the variance of power across stimulus trials divided by the number of trials in each frequency bin. This estimation of neural background noise is more viable for TEMPEST responses than the estimation from neighboring noise bins which is commonly used in case of ASSRs (40). This is because TEMPEST responses are expected to contain evoked power within a certain frequency band whereas ASSRs only have evoked power in the modulation frequency bin. Additionally, as the neural background noise typically exhibits a  $1/f$  spectrum, noise power at the lower frequency side is higher than at the

higher frequency side. Evoked responses to a repeated stimulus expected to be consistent in power and phase across trials, while neural background noise adds a random amplitude and phase to that of the evoked response in each trial. Under this assumption, variance in power across trials divided by the number of trials reflects neural background noise power (61). The two power-based metrics are ubiquitously used in the neuroscience field to indicate the strength and quality of the measured averaged response. The other two metrics are solely based on the phase of the individual response trials: inter-trial phase coherence (ITPC) and pairwise phase consistency (PPC). The first metric, ITPC, indicates consistency of phase-locking to a stimulus based on the magnitude of the average of unit vectors rotated by extracted phases  $\theta_n$  across  $N$  trials (17, 62).

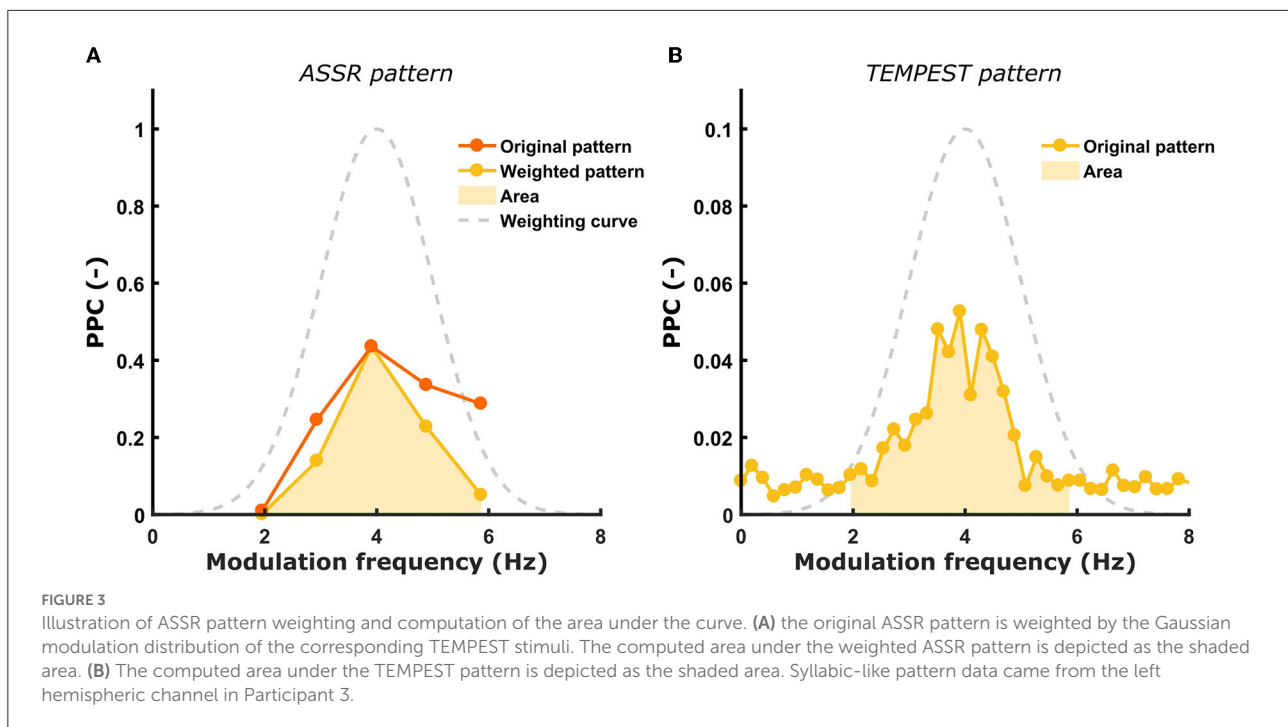
$$ITPC = \left( \sum_{n=1}^N \frac{\cos\theta_n}{N} \right)^2 + \left( \sum_{n=1}^N \frac{\sin\theta_n}{N} \right)^2 \quad (1)$$

The ITPC is commonly used to investigate robustness of phase-locking with different stimulus parameters (17, 18, 55, 57, 62–64). However, despite its considerable presence in the literature, the ITPC is biased by the number of trials with fewer trials resulting in a larger positive bias in the outcome. This bias could hamper comparison between conditions and/or studies with different amounts of trials (58, 61, 65). In contrast to ITPC, the PPC is an unbiased estimate of phase-locking because it is based on the averaged dot product of all possible phase pairs  $\theta_n$  and  $\theta_m$  across  $N$  trials (66).

$$PPC = \frac{2}{N(N-1)} \sum_{n=1}^{N-1} \sum_{m=n+1}^N \cos(\theta_n - \theta_m) \quad (2)$$

When phase consistency is high, then distances between phase pairs will become smaller and thus dot products will be larger. The advantage of the PPC is that it allows for comparison between studies and conditions even with different trial numbers. Both ITPC and PPC take up values between 0 and 1, with 0 indicating no phase-locking at all and 1 indicating perfect phase-locking across trials. Note that ITPC and PPC for TEMPEST responses are computed for each time and frequency bin. In order to obtain electrophysiological patterns as a function of modulation frequency in each participant, results of each metric were averaged in the time domain.

Responses were tested for significance against the neural background noise using the Hotelling  $T^2$  test (52, 67). ASSRs were tested only at their modulation frequency bin while TEMPEST neural activity was tested in each modulation frequency bin of the spectral domain. To evaluate similarity between ASSR patterns and between TEMPEST patterns measured with different electrophysiological metrics and whether different metrics would reveal different characteristics



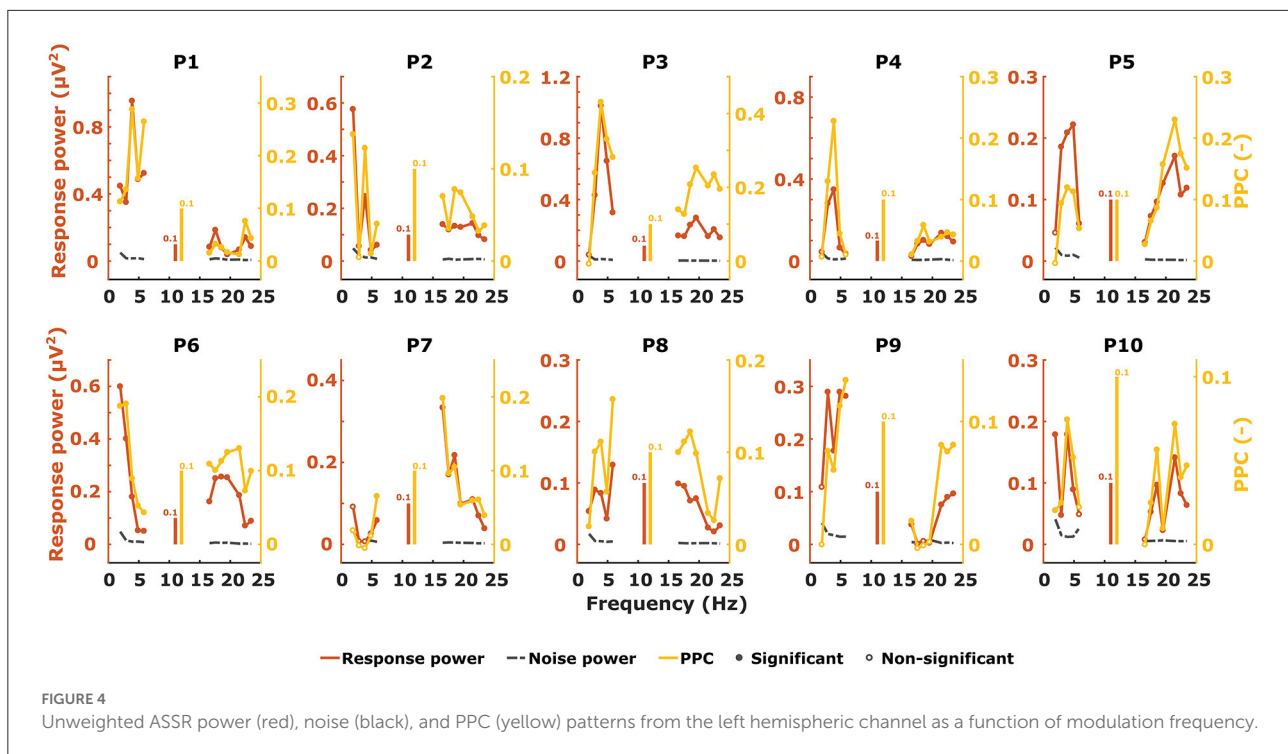
of the neural patterns, the patterns were subjected to correlation analyses. Only significant response bins were included in the analyses. Additionally, because ITPC and PPC are bounded between 0 and 1, their values were first transformed using the Fischer z-transformation. Pearson's correlation coefficients and corresponding  $p$ -values were then reported. The significance level was  $\alpha = 0.05$  at all times and *post-hoc* Bonferroni correction was used to control for false discovery rate since multiple correlations were being tested simultaneously.

### Comparison between TEMPEST and ASSR TMTF patterns

The main goal of this study is to investigate whether the neural activity evoked by TEMPEST stimuli is comparable to the overall ASSR activity (i.e., the TMTF) in the same frequency band. Usually, the TMTF is obtained by setting out ASSR amplitude as a function of modulation frequency (36–39). However, in this study not only ASSR patterns of power, but also of SNR, ITPC and PPC were used. When the ASSR TMTF shows a prominent peak, we hypothesize that the TEMPEST neural activity would also show a relatively large peak and vice versa for each metric. The presence of prominent peaks translates into a larger area under the TMTF pattern. To compare the TEMPEST patterns with those of ASSRs, areas under patterns of the same metric were computed and correlated with each other across all participants in the left and right hemispheres. Before computing the area of ASSR patterns, patterns were first weighted according to the corresponding TEMPEST modulation

distribution in order to account for the relative contribution of each modulation frequency to the TEMPEST neural activity. Each modulation frequency of the SAM stimuli contribute equally to the ASSR patterns. However, these contributions are not equal anymore in case of TEMPEST due to the a-priori modulation frequency distribution used to generate the stimuli. To achieve this weighting of the ASSR pattern, it is multiplied with the Gaussian curve of the corresponding syllabic-like or phonemic-like TEMPEST modulation distribution. By doing this, the TEMPEST and ASSR neural evoked activity can be directly compared to each other after accounting for the modulation distribution shape. Areas under the patterns were computed between 2 and 6 Hz for syllabic-like responses, and between 17 and 23 Hz for phonemic-like responses (Figure 3).

The area was computed by summing up the values in each frequency bin within the restricted band. Finally, to test the relative correspondence between the ASSR and TEMPEST patterns, Pearson's correlations between the TEMPEST and ASSR areas across participants were computed. Only areas of the same metric from ASSR and TEMPEST analyses were correlated (e.g., the area of ASSR SNR was correlated with the area of TEMPEST SNR). Partial Pearson's correlations were computed between TEMPEST and ASSR power area in order to control for any potential effects of induced power area. Induced power is the power that appears in the EEG in any frequency band while listening to a stimulus. In order to investigate whether neural phase-locking to the TEMPEST and SAM stimuli correspond to each other, the correlation with induced power must be controlled for. The induced power



area in the syllabic-like frequency range was computed from the averaged power spectrum of the distinct phonemic-like TEMPEST stimuli, whereas the induced power area in the phonemic-like frequency range was computed from the distinct syllabic-like TEMPEST stimuli. The significance level for the correlations was  $\alpha = 0.05$  and  $p$ -values were corrected with the Bonferroni procedure.

## Results

### Evaluation of electrophysiological metrics

#### ASSR

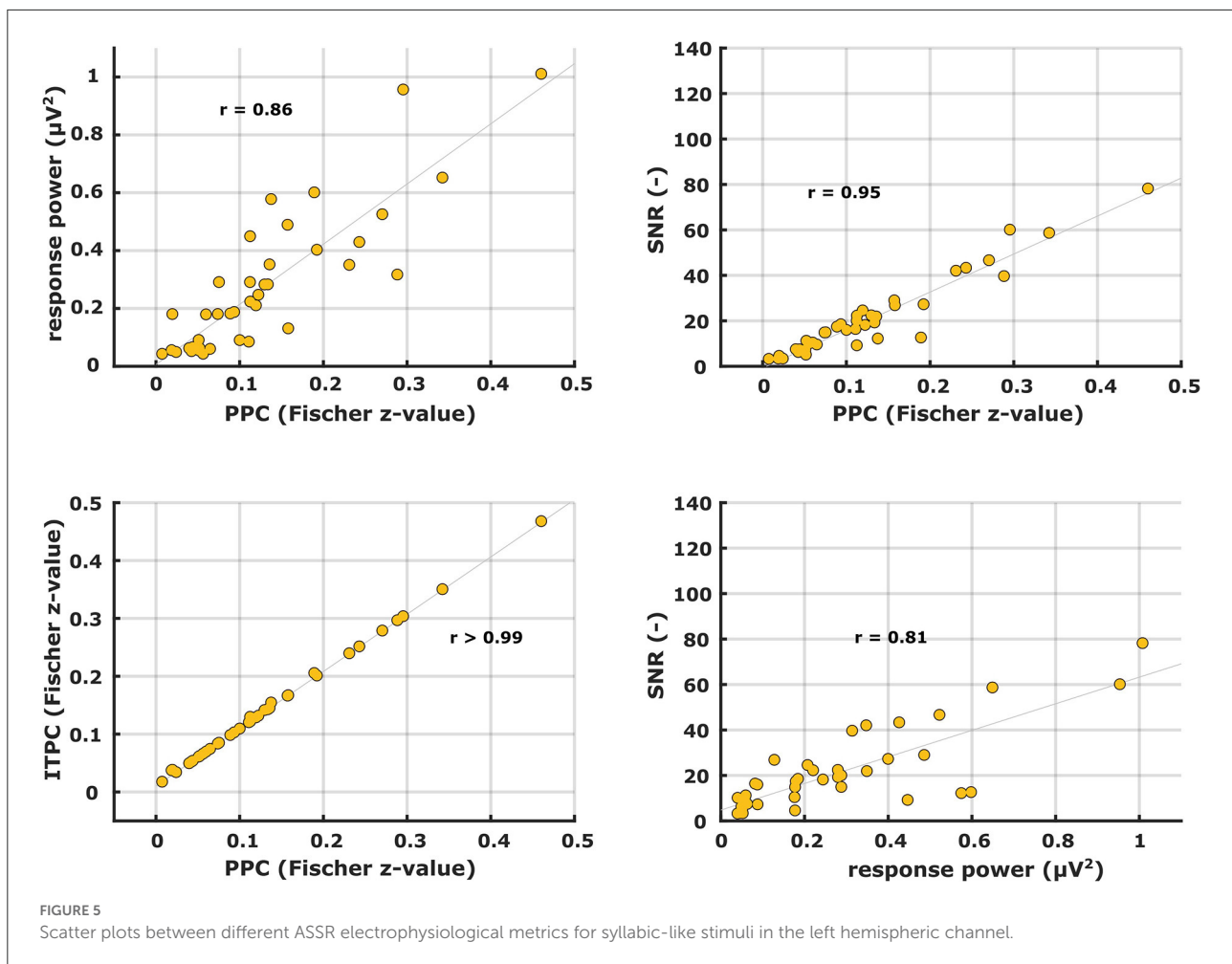
We measured ASSRs with 2–6 Hz (syllabic-like) and 17–23 Hz (phonemic-like) modulation frequencies and obtained the response pattern across modulation frequency for each participant and electrophysiological metric, which is very similar to how TMTFs are obtained elsewhere. Almost all ASSRs were found to be statistically significantly different from noise using the Hotelling  $T^2$  test. Figure 4 shows the individual ASSR patterns measured with response power and PPC for syllabic ASSRs in the left hemispheric channel. In this case, the patterns of these two metrics are relatively similar to each other within each participant. The different shapes of the patterns demonstrate the large variability in ASSRs across modulation frequency and participants.

Patterns of the other electrophysiological metrics are not shown but their similarity in shape to each other was evaluated with correlation analyses. Examples of correlation scatterplots for the syllabic-like responses in the left hemispheric channel are shown in Figure 5. Table 1 summarizes all Pearson's correlation coefficients between the different electrophysiological metrics. Since the response power and PPC patterns were relatively similar, they were highly correlated with each other [ $r(38) = 0.85$ ,  $p < 0.0001$ ]. The ITPC and PPC showed an almost perfect linear correlation (Figure 5, bottom left) based on the fact that the PPC is an unbiased estimate of phase-locking compared to the biased ITPC due to the number of trials. Exchanging the ITPC for PPC would not virtually change the interpretation of the results. The next highest correlations were found between SNR and PPC, which are very high [from  $r(38) = 0.92$  to  $r(64) = 0.99$ ,  $p < 0.0001$ ]. Comparing the ASSR power with SNR and PPC resulted in moderate to high correlation coefficients. Each correlation coefficient was found to be highly significant (Table 1).

#### TEMPEST

When characterizing neural responses to TEMPEST stimuli for each modulation frequency, all electrophysiological metrics showed variation across participants. Figures 6, 7 show only response power and PPC patterns layered over each other for syllabic-like and phonemic-like neural





responses, respectively. Patterns of the other metrics are not shown, but pattern correlations between all four metrics are presented in Table 1. In some participants, distinctive peaks around the mean modulation frequency of the envelope were found in their patterns. For example, participants 1, 3, 5, and 9 showed increased activity around 4 Hz with syllabic-like stimuli. Interestingly, unlike these participants, participant 2 did not show peak activity around 4 Hz but a broader one around 7–8 Hz with syllabic-like stimuli, which corresponds to the range of second harmonic frequencies. With phonemic-like stimuli, participants 3, 5, 6, 7, and 8 showed highly prominent peaks around 20 Hz. As expected, responses to distinct stimuli did not show the increased averaged neural activity as with frozen stimuli.

Patterns of the other metrics are not shown but – like with the ASSRs – their similarity to the PPC patterns was evaluated with correlation analyses. Correlations between the different electrophysiological metric patterns are shown in Table 2. Again, unsurprisingly, the ITPC and PPC showed an almost perfect

linear correlation ( $r > 0.99$ ) (Figure 8, bottom left). Exchanging the ITPC for PPC would not virtually change the interpretation of the results as well in this case. Other metric comparisons resulted in moderate to high correlations except for power vs. PPC in the left hemisphere for phonemic-like stimuli. Correlations with PPC for power and SNR were not as strong as those for ASSRs. Each correlation coefficient was found to be highly significant, except for power vs. PPC in the left hemisphere for phonemic-like stimuli (Table 2).

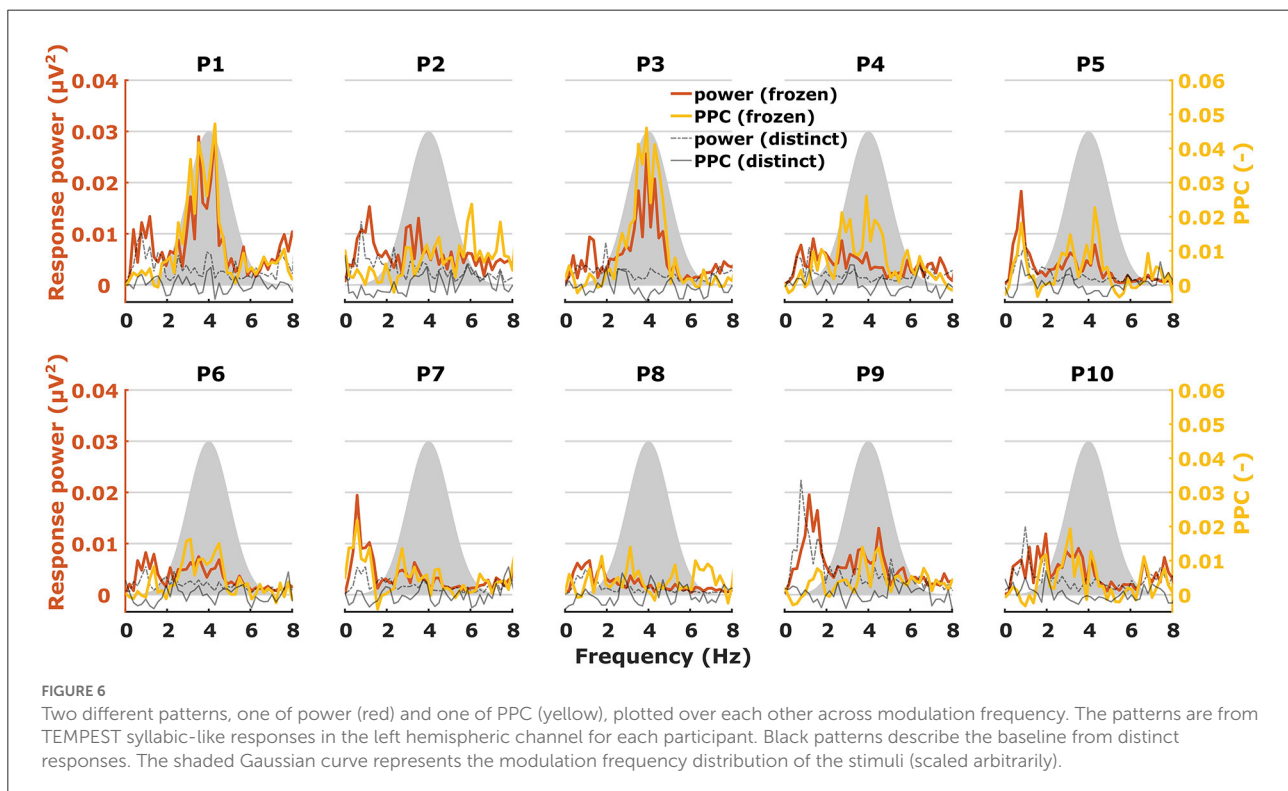
## Comparison of TEMPEST and ASSR neural activity

If an individual ASSR TMTF does not show a prominent peak, then we expect that the TEMPEST neural pattern would not show a peak as well and vice versa. To assess whether the overall activity of ASSR TMTFs corresponds to the activity of TEMPEST responses within participants, we computed the area under the patterns and performed correlation analyses. Only

**TABLE 1** Pearson's correlation coefficients between different electrophysiological metrics of syllabic-like and phonemic-like ASSRs in the left and right hemispheric channels.

	Syllabic-like		Phonemic-like	
	Left ( <i>df</i> = 37)	Right ( <i>df</i> = 40)	Left ( <i>df</i> = 64)	Right ( <i>df</i> = 63)
Power vs. PPC	0.86* (<0.0001)	0.74* (<0.0001)	0.69* (<0.0001)	0.62* (<0.0001)
SNR vs. PPC	0.95* (<0.0001)	0.92* (<0.0001)	0.98* (<0.0001)	0.99* (<0.0001)
ITPC vs. PPC	> 0.99* (<0.0001)	> 0.99* (<0.0001)	> 0.99* (<0.0001)	> 0.99* (<0.0001)
SNR vs. power	0.81* (<0.0001)	0.63* (<0.0001)	0.68* (<0.0001)	0.63* (<0.0001)

Corresponding p-values are reported between parentheses. \*Significant after post-hoc Bonferroni correction.



areas under the ASSR and the TEMPEST pattern of the same electrophysiological metric were used because comparing areas with different metrics would not be insightful (e.g., area under ASSR PPC pattern vs. area under TEMPEST SNR pattern). The computed areas were directly correlated across all ten participants for SNR and PPC of syllabic-like and phonemic-like responses in the left and right hemispheres separately (Figure 9, middle and right columns). Based on the almost perfect correlation between ITPC and PPC (Tables 1, 2), ITPC was left out because it would produce the same results as the PPC. All correlation coefficients were found to be strong [ $r(8) = 0.75-0.98$ ] and highly significant after *post-hoc* Bonferroni correction ( $p \leq 0.001$ ), except for the correlation coefficient between ASSR and TEMPEST SNR area for syllabic-like responses in the right hemispheric channel which was not significant anymore

after *post-hoc* correction ( $p = 0.013$ ). For the power metric, partial Pearson's correlations between TEMPEST power area and ASSR power area were computed in order to control for any potential effects of induced power area. Partial correlation coefficients were found to be strong [ $r(7) = 0.81-0.97$ ] and highly significant ( $p \leq 0.001$ ). These high correlations indicate that the overall activity of the TEMPEST responses corresponds to the speech-weighted overall activity of the ASSR TMTF.

## Discussion

The TEMPEST framework was introduced by Gransier and Wouters (51) to provide an efficient method to investigate the

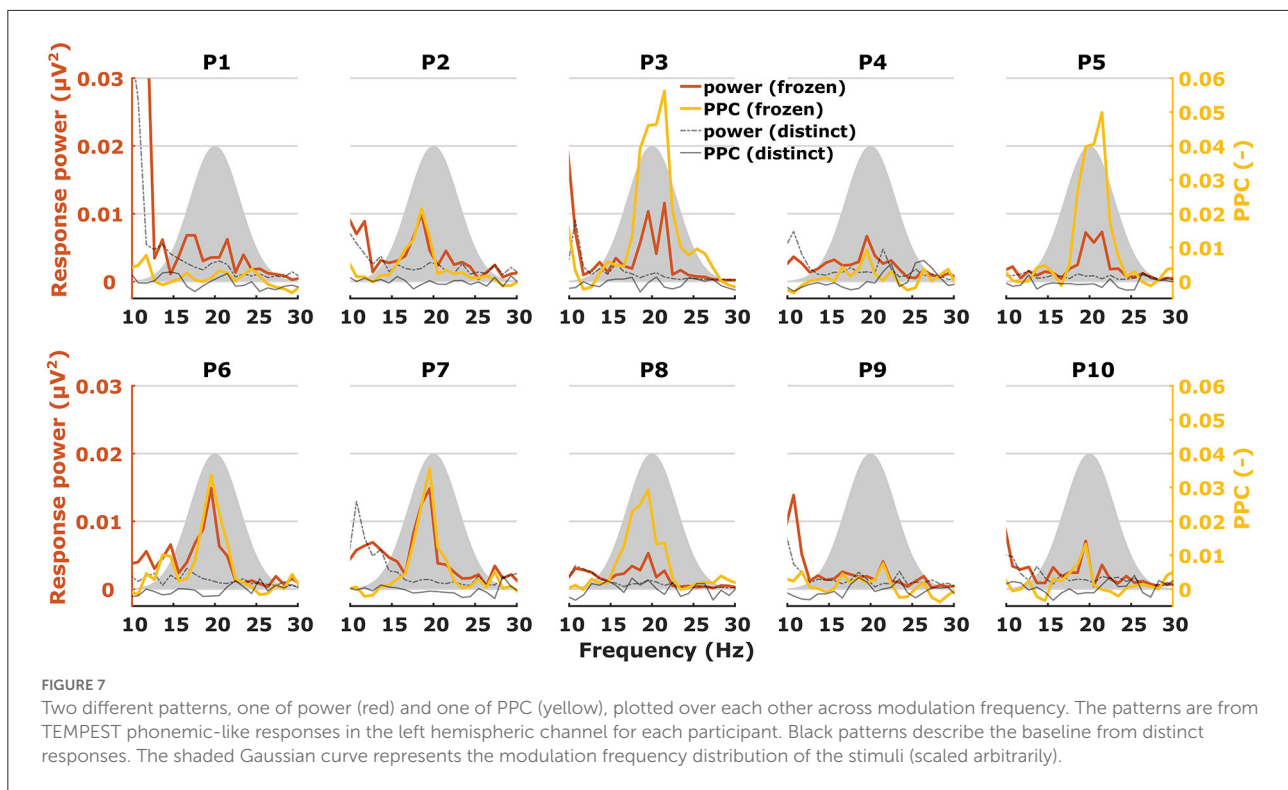


TABLE 2 Pearson's correlation coefficients between different electrophysiological metrics of syllabic-like and phonemic-like TEMPEST neural responses in the left and right hemispheric channels.

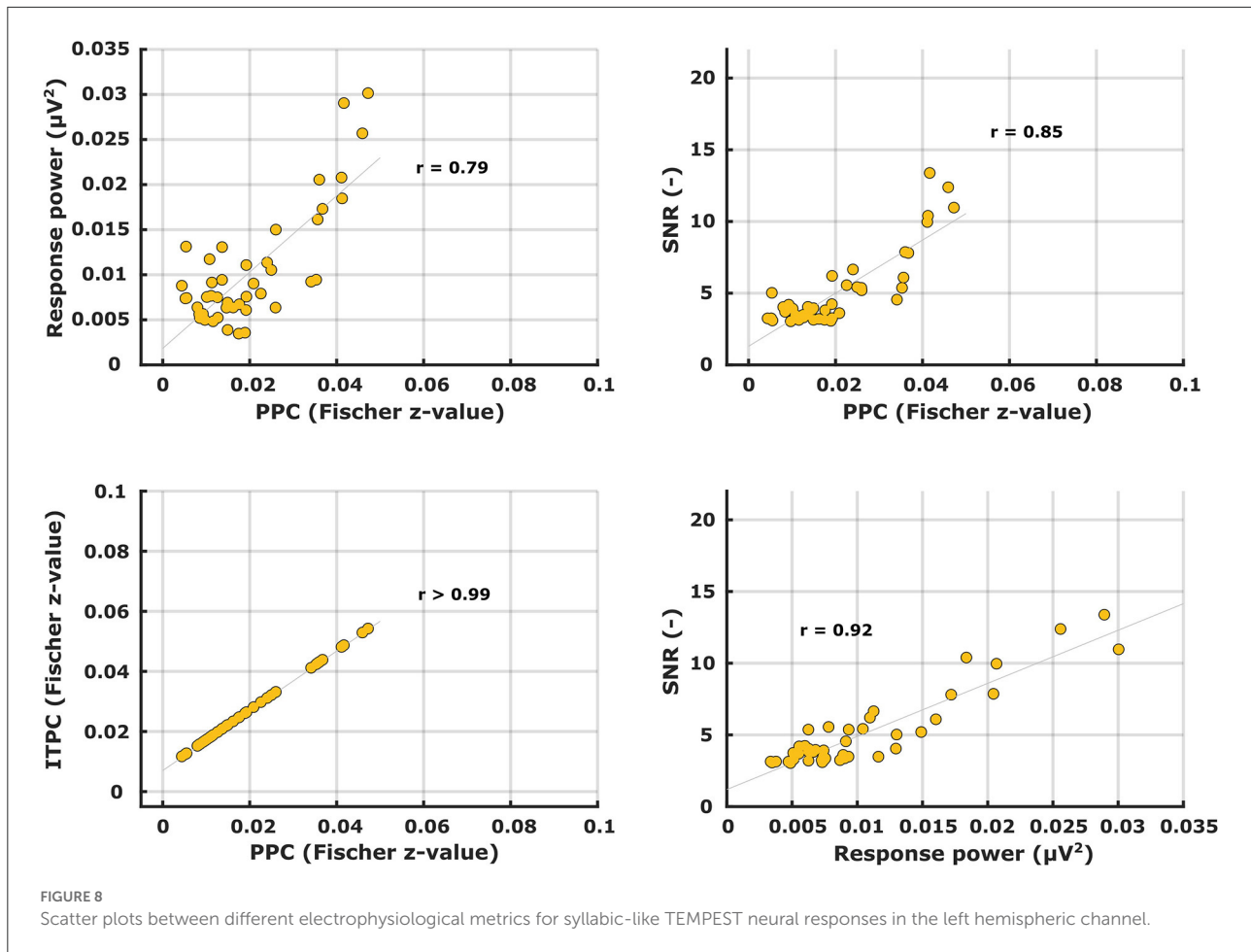
	Syllabic-like		Phonemic-like	
	Left ( <i>df</i> = 42)	Right ( <i>df</i> = 39)	Left ( <i>df</i> = 22)	Right ( <i>df</i> = 27)
Power vs. PPC	0.79* (<0.0001)	0.49* (0.0012)	0.31 (0.14)	0.44 (0.017)
SNR vs. PPC	0.85* (<0.0001)	0.70* (<0.0001)	0.86* (<0.0001)	0.91* (<0.0001)
ITPC vs. PPC	>0.99* (<0.0001)	>0.99* (<0.0001)	>0.99* (<0.0001)	>0.99* (<0.0001)
SNR vs. power	0.92* (<0.0001)	0.70* (<0.0001)	0.49 (0.015)	0.54* (0.0025)

Corresponding p-values are reported between parentheses. \*Significant after post-hoc Bonferroni correction.

neural representation of the stimulus' envelope with speech-like modulation frequencies. In this study, we aimed to demonstrate a proof-of-concept of the TEMPEST framework to efficiently assess the overall capability of temporal envelope encoding in the auditory pathway. To this end, we investigated whether the neural activity evoked by TEMPEST stimuli corresponds to the speech-weighted electrophysiological TMTF, which is classically measured with ASSRs. We used four different electrophysiological metrics to characterize the neural responses. Two metrics were purely based on power (evoked power and SNR) and two other metrics were purely based on phase (ITPC and PPC) of the individual trials or the averaged trial of the neural response. These metrics were computed for each modulation frequency to obtain neural activity patterns

as a function of modulation frequency. This approach is similar to how TMTFs were obtained in other studies using ASSR amplitude (36–39). Comparing the overall neural activity pattern obtained with TEMPEST to the speech-weighted TMTF obtained with ASSRs allowed us to investigate whether they correspond to each other across listeners.

First, we compared neural activity patterns of different metrics with each other for the ASSRs and the TEMPEST responses separately. This is to investigate whether different metrics would reveal different characteristics of the evoked neural activity. A notable case is the almost perfect linear correlation between the ITPC and PPC (Figures 5, 8, bottom left panel) because these two metrics are similar to each other except for a bias due to the number of trials in the ITPC

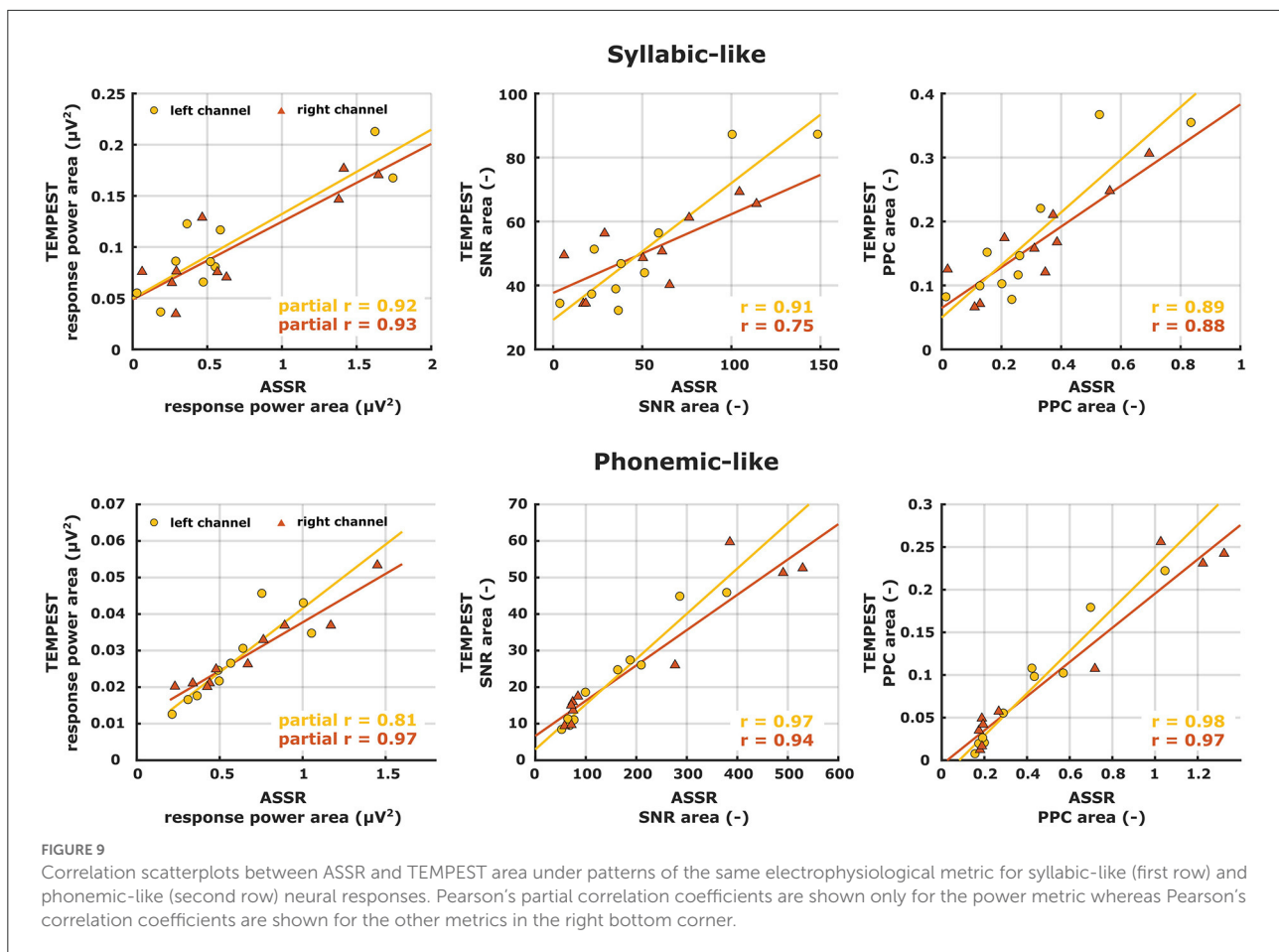


(66). Small deviations occurred because the number of trials was slightly different which resulted in a slightly different bias in ITPC. Furthermore, the bias was relatively small because considerable numbers of trials were used to compute the ITPC (58, 65). Consequently, ITPC results can be exchanged by PPC results without loss of interpretation. As indicated by the high and significant correlation coefficients (Table 1), all individual ASSR TMTF patterns had the same characteristics regardless of the metric used. In the case of TEMPEST activity, syllabic-like patterns of all different metrics significantly correlated with each other. However, phonemic-like patterns of power and PPC did not correlate significantly with each other in both hemispheric channels, and neither did the SNR and power patterns in the left channel (Table 2). Interestingly, the power-based SNR was highly correlated with the phase-based PPC for both ASSRs and TEMPEST responses. While SNR and PPC are based on two independent aspects of the response, i.e., power and phase, the high correlation might be explained by better representation of the phase pattern of the recorded responses due to higher SNR (55, 57). The power metric leads to mostly moderate correlations for both ASSR and TEMPEST stimuli. However, power by itself

doesn't tell much about the presence of a response compared to the presence of background noise, which the SNR and PPC can do to a certain extent. This interaction might explain the smaller correlations between power and the other two metrics. Nevertheless, the high correlations also indicate a high similarity of the intersubject variability in envelope modulation processing across modulation frequency across metrics. For example, if a participant showed a large peak of SNR at a certain modulation frequency, then a large peak of PPC is also expected to appear at the same modulation frequency (Figures 5, 8). Therefore, patterns obtained with different electrophysiological metrics, are comparable.

Patterns obtained with TEMPEST stimuli differed across participants which is consistent with the notion that neural phase-locked activity varies considerably across individuals (36). Participants 1, 3, 5, and 9 had relatively large neural activity peaks around 4 Hz when listening to syllabic-like stimuli, while others showed less prominent or no peaks at all. Participants who had prominent syllabic-like neural activity do not necessarily have prominent phonemic-like neural activity as well (e.g., participant 1 in Figures 6, 7), demonstrating





the variability across modulation frequency as well (36). Interestingly, participant 2 had no prominent neural activity around 4 Hz when listening to syllabic-like stimuli, but it was instead shifted up to around 8 Hz. One likely explanation is that the higher harmonics of the envelope modulations were preferentially encoded and/or processed in the auditory system in this participant, which is more likely for such slow modulation frequencies (68, 69).

Some studies used non-speech stimuli with different irregular envelope characteristics and investigated their evoked response using phase coherence metrics (55, 57). Both studies of Teng and colleagues used stimuli with dynamic acoustic changes that occur at timescales similar to our stimuli. They used several different stimuli with dynamics at different timescales, some of which coincided with those of syllables and phonemes in speech. Two of those stimuli were the theta- and gamma-sounds. The theta-sound contained changes at a mean timescale of 190 ms ( $\sim 5$  Hz modulation rate) which approximately corresponds to the syllable mean modulation frequency of our syllabic-like TEMPEST stimuli. Similarly, the gamma-sound was temporally related to the phoneme rate with

a mean timescale of 27 ms ( $\sim 37$  Hz modulation rate). The authors computed the ITPC of the brain's response for each modulation frequency [note that they used the formula from Lachaux et al. (70), not formula (1) in this study]. Responses evoked by theta sounds showed significantly increased ITPC around 4 Hz and those evoked by gamma sounds around 37 Hz. The peaks that we found in the neural patterns within the modulation frequency range of the TEMPEST stimuli are reminiscent of this finding. Teng and Poeppel (57) also included beta-sounds with mean timescales of 62 and 41 ms (modulation rates of  $\sim 16$  and  $\sim 24$  Hz, respectively), thus these stimuli are temporally more closely related to our phonemic-like stimuli. However, they reported a considerable decrease in ITPC with beta sounds compared to theta and gamma sounds. In contrast, we did not find a decrease in ITPC and PPC with phonemic-like TEMPEST stimuli compared to syllabic-like TEMPEST stimuli, and similar conclusions can also be made in the case of response power (Figures 6, 7). Another study by Teng and colleagues used complex stimuli with irregular 1/f modulation spectra (56). They investigated robustness of neural phase-locking by comparing ITPC results with frozen

and distinct stimuli ( $n = 25$ ). To this end, the ITPC of the distinct stimuli was subtracted from the frozen ITPC. In a way, this is subtracting the bias from the frozen ITPC and this would be comparable to the PPC. The ITPC difference that they found was at approximately 0.06 in the delta and theta band, which is in line with our syllabic-like results (Figure 6).

Luo and colleagues have also looked at the difference in ITPC between responses evoked by the same (frozen) spoken sentence and responses evoked by different (distinct) sentences (17, 62). ITPC differences of responses to spoken sentences in the delta-theta band are comparable to our syllabic-like PPC results. Another study used mutual information to investigate how much the response phase in the theta band encodes information about the sentence stimulus (71). Peaks of mutual information in the theta band varied across participants, which is in line with the variability in ASSR TMTF for low frequencies (36) and with our results that show variable peaks of activity using syllabic-like stimuli. Additionally, small peaks of mutual information were present in the 22–27 Hz range in some participants and were slightly visible in the grand-average pattern. This frequency range is close to the modulation frequency range of our phonemic-like stimuli. Furthermore, the difference in order of magnitude in mutual information between the theta band and the 22–27 Hz range is similar to the difference that our results exhibit between the syllabic-like and phonemic-like responses. This similarity should be treated with caution because our metrics are not related to mutual information. One thing to keep in mind is that sentence stimuli contain a much wider range of modulation frequencies than our syllabic-like and phonemic-like TEMPEST stimuli.

The main goal of the study was to evaluate whether the global neural activity evoked by TEMPEST stimuli was qualitatively comparable to the speech-weighted overall activity in the electrophysiological TMTF measured with ASSRs. To this end, we computed the area under the patterns of power, SNR, and PPC as a function of modulation frequency of TEMPEST responses and area under the ASSR TMTFs by summing up the values at significant response frequency bins. Before the computation of the area, TMTFs were first weighted with the Gaussian curve of the modulation frequency distribution from the corresponding TEMPEST stimuli. We then computed same-metric correlation coefficients between these areas across participants. All correlations between ASSR and TEMPEST were found to be strong and significant except for the SNR in the right hemispheric channel (Figure 9). These significantly high correlations indicate that the neural activity evoked by TEMPEST stimuli is comparable to those of the speech-weighted TMTF measured with the classical ASSR paradigm. Furthermore, they also show that the variability in the global neural patterns across listeners as measured with TEMPEST stimuli is similar to that found with ASSR TMTFs, which is consistent with the findings by

(36). Consequently, evoked TEMPEST responses characterized by any of the three metrics (power, SNR, or PPC) can be used as an indicator of individual neural temporal processing capability within the modulation frequency band of interest.

Although our approach of computing the area under the patterns of TEMPEST neural activity and the TMTF does not consider the exact pattern shapes, we found that the overall activity evoked by TEMPEST stimuli strongly corresponds to the overall activity found in the electrophysiological TMTF. This result is a clear indication that the TEMPEST framework has the potential to evaluate temporal envelope processing in the auditory pathway. Furthermore, since TEMPEST stimuli contain a range of envelope modulations as determined by an a-priori modulation frequency distribution, individual distribution-weighted electrophysiological TMTFs can be efficiently determined, which would otherwise be measured by multiple SAM stimuli, as is clear from Figures 6, 7. Further research on variations of TEMPEST stimuli and improvement of the neurophysiological analyses can potentially push the TEMPEST framework to more clinical usability. Moreover, the TEMPEST framework provides many possibilities to generate TEMPEST stimuli that are parameterized, for example, by a modulation frequency distribution, a modulation depth distribution, window shape with optionally varying parameters, etc. Furthermore, the framework also allows for more complex stimuli such as nesting of two or more TEMPEST envelopes (51), which combines multiple TEMPEST stimuli with different modulation frequency distributions into one stimulus. This approach would be comparable to combining multiple SAM stimuli at different carrier frequencies and is commonly used to electrophysiologically determine frequency-specific hearing thresholds in infants (72).

## Conclusion

The TEMPEST framework (51) provides stimuli that evoke neural phase-locked activity with the same characteristics as the electrophysiological TMTF classically measured with ASSRs after weighting by the TEMPEST distribution. Since TEMPEST stimuli contain a range of envelope modulation frequencies in contrast to single-frequency SAM stimuli, they can be used to efficiently probe temporal envelope processing in the auditory pathway. Any of the four electrophysiological metrics (evoked power, SNR, ITPC, or PPC) can be used to evaluate the degree of neural tracking to amplitude-modulated stimuli. Moreover, TEMPEST stimuli that contain speech-like modulations (such as the syllable and the phoneme rate in speech) have the potential to provide a better understanding of the role of neural envelope processing in speech perception. Not only that, but they could also

potentially capture differences in temporal envelope processing in different listener groups with different types of auditory processing deficits. Future work would further investigate the potential of the TEMPEST framework using more complex stimuli by varying several other envelope parameters or combining different stimuli into one stimulus with multiple bands of modulation frequencies, and explore different analysis techniques to exploit its full potential in the neuroscientific and audiological fields.

## Data availability statement

The data generated and/or analyzed during the current study are not publicly available for legal/ethical reasons, they can be obtained on reasonable request to the corresponding author.

## Ethics statement

The studies involving human participants were reviewed and approved by Medical Ethical Committee of the University Hospitals and University of Leuven. The patients/participants provided their written informed consent to participate in this study.

## Author contributions

All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

## References

- Plomp R. The role of modulation in hearing. In: Klinke R, Hartmann R, editors. *HEARING—Physiological Bases and Psychophysics*. Berlin: Springer (1983).
- Rosen S. Temporal information in speech: acoustic, auditory and linguistic aspects. *Philos Trans R Soc Lond B Biol Sci*. (1992) 336:367–73. doi: 10.1098/rstb.1992.0070
- Ding N, Patel AD, Chen L, Butler H, Luo C, Poeppel D. Temporal modulations in speech and music. *Neurosci Biobehav Rev*. (2017) 81:181–7. doi: 10.1016/j.neubiorev.2017.02.011
- Varnet L, Ortiz-Barajas MC, Erra RG, Gervain J, Lorenzi C. A cross-linguistic study of speech modulation spectra. *J Acoust Soc Am*. (2017) 142:1976–89. doi: 10.1121/1.5006179
- Greenberg S, Carvey H, Hitchcock L, Chang S. Temporal properties of spontaneous speech - a syllable-centric perspective. *J Phon*. (2003) 31:465–85. doi: 10.1016/j.wocn.2003.09.005
- Goswami U, Leong V. Speech rhythm and temporal structure: converging perspectives? *Lab Phonol*. (2013) 4:67–92. doi: 10.1515/lp-2013-0004
- Greenberg S. Speaking in shorthand - a syllable-centric perspective for understanding pronunciation variation. *Speech Commun*. (1999) 29:159–76. doi: 10.1016/S0167-6393(99)00050-3
- Drullman R, Festen JM, Plomp R. Effect of reducing slow temporal modulations on speech reception. *J Acoust Soc Am*. (1994) 95:2670–80. doi: 10.1121/1.409836

## Funding

This work was partly funded by a research grant from Flanders Innovation & Entrepreneurship through the VLAIO research grant HBC.20192373, partly by a Wellcome Trust Collaborative Award in Science RG91976 to Robert P. Carlyon, John C. Middlebrooks, and JW, and partly by an SB Ph.D. grant 1S34121N from the Research Foundation Flanders (FWO) awarded to WD.

## Acknowledgments

We are very grateful to all participants for their help in this study.

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

- Shannon R V, Zeng F-G, Kamath V, Wygonski J, Ekelid M. Speech recognition with primarily temporal cues. *Science*. (1995) 270:303–4. doi: 10.1126/science.270.5234.303
- Smith ZM, Delgutte B, Oxenham AJ. Chimaeric sounds reveal dichotomies in auditory perception. *Nature*. (2002) 416:87–90. doi: 10.1038/416087a
- Zeng FG, Nie K, Stickney GS, Kong YY, Vongphoe M, Bhargava A, et al. Speech recognition with amplitude and frequency modulations. *Proc Natl Acad Sci U S A*. (2005) 102:2293–8. doi: 10.1073/pnas.0406460102
- Friesen LM, Shannon R V, Baskent D, Wang X. Speech recognition in noise as a function of the number of spectral channels: comparison of acoustic hearing and cochlear implants. *J Acoust Soc Am*. (2001) 110:1150–63. doi: 10.1121/1.1381538
- Peelle JE, Davis MH. Neural oscillations carry speech rhythm through to comprehension. *Front Psychol*. (2012) 3:1–17. doi: 10.3389/fpsyg.2012.00320
- Abrams DA, Nicol T, Zecker S, Kraus N. Right-hemisphere auditory cortex is dominant for coding syllable patterns in speech. *J Neurosci*. (2008) 28:3958–65. doi: 10.1523/JNEUROSCI.0187-08.2008
- Aiken SJ, Picton TW. Human cortical responses to the speech envelope. *Ear Hear*. (2008) 29:139–57. doi: 10.1097/AUD.0b013e31816453dc
- Ahissar E, Nagarajan S, Ahissar M, Protopapas A, Mahncke H, Merzenich MM. Speech comprehension is correlated with temporal response patterns recorded from auditory cortex. *Proc Natl Acad Sci U S A*. (2001) 98:13367–72. doi: 10.1073/pnas.201400998

17. Luo H, Poeppel D. Phase patterns of neuronal responses reliably discriminate speech in human auditory cortex. *Neuron*. (2007) 54:1001–10. doi: 10.1016/j.neuron.2007.06.004
18. Howard MF, Poeppel D. Discrimination of speech stimuli based on neuronal response phase patterns depends on acoustics but not comprehension. *J Neurophysiol*. (2010) 104:2500–11. doi: 10.1152/jn.00251.2010
19. Vanthornhout J, Decruy L, Wouters J, Simon JZ, Francart T. Speech intelligibility predicted from neural entrainment of the speech envelope. *J Assoc Res Otolaryngol*. (2018) 19:181–91. doi: 10.1007/s10162-018-0654-z
20. Ding N, Simon JZ. Adaptive temporal encoding leads to a background-insensitive cortical representation of speech. *J Neurosci*. (2013) 33:5728–35. doi: 10.1523/JNEUROSCI.5297-12.2013
21. Decruy L, Vanthornhout J, Francart T. Evidence for enhanced neural tracking of the speech envelope underlying age-related speech-in-noise difficulties. *J Neurophysiol*. (2019) 122:601–15. doi: 10.1152/jn.00687.2018
22. Riecke L, Formisano E, Sorger B, Başkent D, Gaudrain E. Neural entrainment to speech modulates speech intelligibility. *Curr Biol*. (2018) 28:161–9.e5. doi: 10.1016/j.cub.2017.11.033
23. Di Liberto GM, O'Sullivan JA, Lalor EC. Low-frequency cortical entrainment to speech reflects phoneme-level processing. *Curr Biol*. (2015) 25:2457–65. doi: 10.1016/j.cub.2015.08.030
24. Peelle JE, Gross J, Davis MH. Phase-locked responses to speech in human auditory cortex are enhanced during comprehension. *Cereb Cortex*. (2013) 23:1378–87. doi: 10.1093/cercor/bhs118
25. Gross J, Hoogenboom N, Thut G, Schyns P, Panzeri S, Belin P, et al. Speech rhythms and multiplexed oscillatory sensory coding in the human brain. *PLoS Biol*. (2013) 11:e1001752. doi: 10.1371/journal.pbio.1001752
26. Molinaro N, Lizarazu M. Delta (but not theta)-band cortical entrainment involves speech-specific processing. *Eur J Neurosci*. (2018) 48:2642–50. doi: 10.1111/ejn.13811
27. Bonhage CE, Meyer L, Gruber T, Friederici AD, Mueller JL. Oscillatory EEG dynamics underlying automatic chunking during sentence processing. *Neuroimage*. (2017) 152:647–57. doi: 10.1016/j.neuroimage.2017.03.018
28. Etard O, Reichenbach T. Neural speech tracking in the theta and in the delta frequency band differentially encode clarity and comprehension of speech in noise. *J Neurosci*. (2019) 39:5750–9. doi: 10.1523/JNEUROSCI.1828-18.2019
29. Ding N, Melloni L, Zhang H, Tian X, Poeppel D. Cortical tracking of hierarchical linguistic structures in connected speech. *Nat Neurosci*. (2016) 19:158–64. doi: 10.1038/nn.4186
30. Getz H, Ding N, Newport EL, Poeppel D. Cortical tracking of constituent structure in language acquisition. *Cognition*. (2018) 181:135–40. doi: 10.1016/j.cognition.2018.08.019
31. Houtgast T, Steeneken HJM. A review of the MTF concept in room acoustics and its use for estimating speech intelligibility in auditoria. *J Acoust Soc Am*. (1985) 77:1069–77. doi: 10.1121/1.392224
32. Obleser J, Herrmann B, Henry MJ. Neural oscillations in speech: don't be enslaved by the envelope. *Front Hum Neurosci*. (2012) 6:2008–11. doi: 10.3389/fnhum.2012.00250
33. Ding N, Simon JZ. Cortical entrainment to continuous speech: functional roles and interpretations. *Front Hum Neurosci*. (2014) 8:1–7. doi: 10.3389/fnhum.2014.00311
34. Zoefel B, Vanrullen R. The role of high-level processes for oscillatory phase entrainment to speech sound. *Front Hum Neurosci*. (2015) 9:1–12. doi: 10.3389/fnhum.2015.00651
35. Ramus F, Nespore M, Mehler J. Correlates of linguistic rhythm in the speech signal. *Cognition*. (2000) 75:265–92. doi: 10.1016/S0010-0277(99)00058-X
36. Gransier R, Hofmann M, Wieringen A Van, Wouters J. Stimulus-evoked phase-locked activity along the human auditory pathway strongly varies across individuals. *Sci Rep*. (2021) 11:143. doi: 10.1038/s41598-020-80229-w
37. Purcell DW, John SM, Schneider BA, Picton TW. Human temporal auditory acuity as assessed by envelope following responses. *J Acoust Soc Am*. (2004) 116:3581–93. doi: 10.1121/1.1798354
38. Poulsen C, Picton TW, Paus T. Age-related changes in transient and oscillatory brain responses to auditory stimulation during early adolescence. *Dev Sci*. (2009) 12:220–35. doi: 10.1111/j.1467-7687.2008.00760.x
39. Ross B, Borgmann C, Draganova R, Roberts LE, Pantev C. A high-precision magnetoencephalographic study of human auditory steady-state responses to amplitude-modulated tones. *J Acoust Soc Am*. (2000) 108:679–91. doi: 10.1121/1.429600
40. Picton TW, John MS, Dimitrijevic A, Purcell D. Human auditory steady-state responses. *Int J Audiol*. (2003) 42:177–219. doi: 10.3109/14992020309101316
41. Darestani EF, Goossens T, Wouters J, van Wieringen A. Spatiotemporal reconstruction of auditory steady-state responses to acoustic amplitude modulations: Potential sources beyond the auditory pathway. *Neuroimage*. (2017) 148:240–53. doi: 10.1016/j.neuroimage.2017.01.032
42. Herdman AT, Lins O, Van Roon P, Stapells DR, Scherg M, Picton TW. Intracerebral sources of human auditory steady-state responses. *Brain Topogr*. (2002) 15:69–86. doi: 10.1023/A:1021470822922
43. Luke R, De Vos A, Wouters J. Source analysis of auditory steady-state responses in acoustic and electric hearing. *Neuroimage*. (2017) 147:568–76. doi: 10.1016/j.neuroimage.2016.11.023
44. Bidelman GM. Multichannel recordings of the human brainstem frequency-following response: Scalp topography, source generators, and distinctions from the transient ABR. *Hear Res*. (2015) 323:68–80. doi: 10.1016/j.heares.2015.01.011
45. Gransier R, Luke R, Van Wieringen A, Wouters J. Neural modulation transmission is a marker for speech perception in noise in cochlear implant users. *Ear Hear*. (2020) 41:591–602. doi: 10.1097/AUD.0000000000000783
46. Leigh-Paffenroth ED, Fowler CG. Amplitude-modulated auditory steady-state responses in younger and older listeners. *J Am Acad Audiol*. (2006) 17:582–97. doi: 10.3766/jaaa.17.8.5
47. Dimitrijevic A, John MS, Picton TW. Auditory steady-state responses and word recognition scores in normal-hearing and hearing-impaired adults. *Ear Hear*. (2004) 25:68–84. doi: 10.1097/01.AUD.0000111545.71693.48
48. Goossens T, Vercammen C, Wouters J, van Wieringen A. Neural envelope encoding predicts speech perception performance for normal-hearing and hearing-impaired adults. *Hear Res*. (2018) 370:189–200. doi: 10.1016/j.heares.2018.07.012
49. Poelmans H, Luts H, Vandermosten M, Boets B, Ghesquière P, Wouters J. Auditory steady state cortical responses indicate deviant phonemic-rate processing in adults with dyslexia. *Ear Hear*. (2012) 33:134–43. doi: 10.1097/AUD.0b013e31822c26b9
50. Alaerts J, Luts H, Hofmann M, Wouters J. Cortical auditory steady-state responses to low modulation rates. *Int J Audiol*. (2009) 48:582–93. doi: 10.1080/14992020902894558
51. Gransier R, Wouters J. Neural auditory processing of parameterized speech envelopes. *Hear Res*. (2021) 412:108374. doi: 10.1016/j.heares.2021.108374
52. Hofmann M, Wouters J. Improved electrically evoked auditory steady-state response thresholds in humans. *J Assoc Res Otolaryngol*. (2012) 13:573–89. doi: 10.1007/s10162-012-0321-8
53. Van Wieringen A, Wouters J. LIST and LINT: sentences and numbers for quantifying speech understanding in severely impaired listeners for Flanders and the Netherlands. *Int J Audiol*. (2008) 47:348–55. doi: 10.1080/14992020801895144
54. Crystal TH, House AS. Segmental durations in connected-speech signals: current results. *J Acoust Soc Am*. (1988) 83:1553–73. doi: 10.1121/1.395911
55. Teng X, Tian X, Rowland J, Poeppel D. Concurrent temporal channels for auditory processing: Oscillatory neural entrainment reveals segregation of function at different scales. *PLoS Biol*. (2017) 15:1–29. doi: 10.1371/journal.pbio.2000812
56. Teng X, Tian X, Doelling K, Poeppel D. Theta band oscillations reflect more than entrainment: behavioral and neural evidence demonstrates an active chunking process. *Eur J Neurosci*. (2018) 48:2770–82. doi: 10.1111/ejn.13742
57. Teng X, Poeppel D. Theta and gamma bands encode acoustic dynamics over wide-ranging timescales. *Cereb Cortex*. (2020) 30:2600–14. doi: 10.1093/cercor/bhz263
58. Bastos AM, Schoffelen JM. A tutorial review of functional connectivity analysis methods and their interpretational pitfalls. *Front Syst Neurosci*. (2016) 9:1–23. doi: 10.3389/fnsys.2015.00175
59. Jasper HH. The ten twenty electrode system of the international federation. *Electroencephalogr Clin Neurophysiol*. (1957) 10:371–5.
60. Gransier R, van Wieringen A, Wouters J. Binaural interaction effects of 30–50 Hz auditory steady state responses. *Ear Hear*. (2017) 38:e305–15. doi: 10.1097/AUD.0000000000000429
61. Cohen MX. *Analyzing Neural Time Series Data*. MIT Press (2014).
62. Luo H, Liu Z, Poeppel D. Auditory cortex tracks both auditory and visual stimulus dynamics using low-frequency neuronal phase modulation. *PLoS Biol*. (2010) 8:25–6. doi: 10.1371/journal.pbio.1000445



63. Kayser C, Wilson C, Safaai H, Sakata S, Panzeri S. Rhythmic auditory cortex activity at multiple timescales shapes stimulus–response gain and background firing. *J Neurosci.* (2015) 35:7750–62. doi: 10.1523/JNEUROSCI.0268-15.2015
64. VanRullen R. How to evaluate phase differences between trial groups in ongoing electrophysiological signals. *Front Neurosci.* (2016) 10:1–22. doi: 10.3389/fnins.2016.00426
65. Van Diepen RM, Mazaheri A. The caveats of observing inter-trial phase-coherence in cognitive neuroscience. *Sci Rep.* (2018) 8:1–9. doi: 10.1038/s41598-018-20423-z
66. Vinck M, van Wingerden M, Womelsdorf T, Fries P, Pennartz CMA. The pairwise phase consistency: a bias-free measure of rhythmic neuronal synchronization. *Neuroimage.* (2010) 51:112–22. doi: 10.1016/j.neuroimage.2010.01.073
67. Hotelling H. The generalization of student's ratio. *Ann Math Stat.* (1931) 2:360–78. doi: 10.1214/aoms/1177732979
68. Tlumak AI, Durrant JD, Delgado RE, Boston JR. Steady-state analysis of auditory evoked potentials over a wide range of stimulus repetition rates: profile in adults. *Int J Audiol.* (2011) 50:448–58. doi: 10.3109/14992027.2011.560903
69. Tlumak AI, Durrant JD, Delgado RE, Boston JR. Steady-state analysis of auditory evoked potentials over a wide range of stimulus repetition rates: Profile in children vs. adults. *Int J Audiol.* (2012) 51:480–90. doi: 10.3109/14992027.2012.664289
70. Lachaux J-P, Rodriguez E, Martinerie J, Varela FJ. Measuring phase synchrony in brain signals. *Hum Brain Mapp.* (1999) 8:194–208. doi: 10.1002/(SICI)1097-0193(1999)8:4<194::AID-HBM4>3.0.CO;2-C
71. Cogan GB, Poeppel D. A mutual information analysis of neural coding of speech by low-frequency MEG phase information. *J Neurophysiol.* (2011) 106:554–563. doi: 10.1152/jn.00075.2011
72. John MS, Picton TW, MASTER. A windows program for recording multiple auditory steady-state responses. *Comput Methods Programs Biomed.* (2000) 61:125–50. doi: 10.1016/S0169-2607(99)00035-8