Sustained neural rhythms reveal endogenous oscillations supporting speech perception

Authors: Sander van Bree[a,b], Ediz Sohoglu[a,c], Matthew H Davis[a,f], Benedikt Zoefel[a,d,e,f,*]

Affiliations: [a] MRC Cognition and Brain Sciences Unit, University of Cambridge, 15 Chaucer Road, Cambridge CB27EF, UK

[b] School of Psychology and Centre for Human Brain Health, University of Birmingham, Birmingham B15 2TT, UK

[c] School of Psychology, University of Sussex, Pevensey Building, Brighton, BN1 9QH, UK

[d] Centre de Recherche Cerveau et Cognition (CerCo), CNRS UMR 5549, CHU Purpan, Pavillon Baudot, 31052 Toulouse, France

[e] Université Toulouse III Paul Sabatier, Toulouse, France

[f] senior author

*Corresponding author (benedikt.zoefel@cnrs.fr)

Abstract

Rhythmic sensory or electrical stimulation will produce rhythmic brain responses. These rhythmic responses are often interpreted as endogenous neural oscillations aligned to the stimulus rhythm. However, stimulus-aligned brain responses can also be explained as a sequence of evoked responses, which only appear regular due to the rhythmicity of the stimulus, without necessarily involving underlying neural oscillations. To distinguish evoked responses from true oscillatory activity, we tested whether rhythmic stimulation produces oscillatory responses which continue after the end of the stimulus. Such sustained effects provide evidence for true involvement of neural oscillations. In Experiment 1, we found that rhythmic intelligible, but not unintelligible speech produces oscillatory responses in magnetoencephalography (MEG) which outlast the stimulus at parietal sensors. In Experiment 2, we found that transcranial alternating current stimulation (tACS) leads to rhythmic fluctuations in speech perception outcomes which continue after the end of electrical stimulation. We further report that the phase relation between electroencephalography (EEG) and rhythmic intelligible speech can predict the tACS phase that leads to most accurate speech perception. Together, our results lay the foundation for a new account of speech perception which includes endogenous neural oscillations as a key underlying principle.

Introduction

The alignment of oscillatory neural activity to a rhythmic stimulus, often termed "neural entrainment", is an integral part of many current theories of speech processing (1–4). Indeed, brain responses seem to align more reliably to intelligible than to unintelligible speech (5, 6). Similarly, rhythmic electrical stimulation applied to the scalp (tACS) is assumed to "entrain" brain oscillations and has been shown to modulate speech processing and perception (7–11). Despite the prominence of entrainment theories in speech research and elsewhere (1, 12–14), it has been surprisingly difficult to demonstrate that stimulus-aligned brain responses indeed involve endogenous neural oscillations. This is because, if each stimulus in a rhythmic sequence produces a brain response, the evoked brain responses will appear rhythmic as well, without necessarily involving endogenous neural oscillations. This is not only true for sensory stimulation: Rhythmic behavioural effects of tACS cannot be interpreted as evidence of entrained endogenous oscillations; they might simply reflect the impact of regular changes in current imposed onto the brain (15).

In the present work, we provide evidence that rhythmic intelligible speech and tACS entrain endogenous neural oscillations. Neural oscillations are often proposed to align their high-excitability phase to important events in a rhythmic sequence so as to boost the processing of these events and enhance corresponding task performance (12, 13). It is possible that such a process entails a passive, "bottom-up" component during which oscillations are rhythmically "pushed" by the stimulus, similar to the regular swing of a pendulum. On the other hand (and not mutually exclusive), an active, "top-down" component could adjust neural activity so that it is optimally aligned with a predicted stimulus. Importantly, in both cases we would anticipate that oscillatory brain responses are sustained for some time after the offset of stimulation: This could be because predictions about upcoming rhythmic input are upheld, and/or neural oscillations are self-sustaining and (much like a pendulum swing) will continue after the cessation of a driving input. Consequently, sustained oscillatory responses produced by a rhythmic stimulus *after* the cessation of that stimulus can provide evidence for entrainment of endogenous neural oscillations (16, 17).

In this paper, we will contrast this theory of entrained oscillations with an alternative view in which entrainment is merely due to responses evoked directly by the stimulus *per se*. Note that both views are sufficient to accommodate existing evidence of brain signals aligned to a stimulus while the latter is present. Given the difficulty of distinguishing true oscillations from other responses during rhythmic input, we use the term "entrained" only to describe a signal aligned to a stimulus (irrespective of whether this alignment reflects oscillations or evoked responses; see "entrainment in the broad sense" in (14)). We then measure sustained rhythmic activity to infer its neural origins: Truly oscillatory activity that was entrained to the rhythmic stimulus would lead to sustained rhythmic responses, but sustained responses would not be expected for stimulus-evoked neural activity. In the current study, we provide

two distinct sources of evidence for sustained oscillatory effects: (1) oscillatory MEG responses that continue after rhythmic intelligible speech and (2) oscillatory effects of tACS on speech perception that continue after the termination of electrical stimulation. Furthermore, we link these two effects in single participants to show how the phase of oscillatory neural responses measured with EEG can predict the tACS phase at which word report is enhanced. In combination, these findings provide evidence that endogenous neural oscillations in entrained brain responses play a causal role in supporting speech perception.

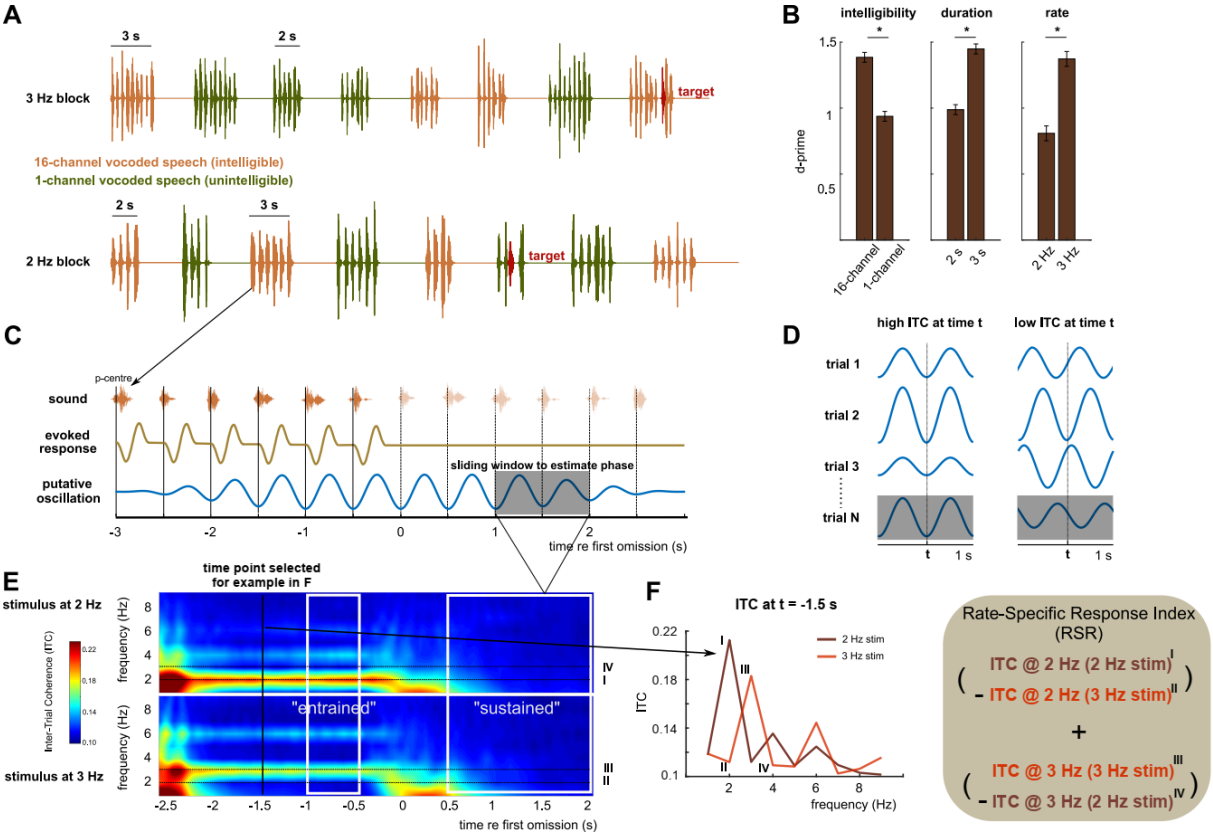

Figure 1. Experimental paradigm and analysis. A. Participants listened to rhythmic speech sequences which differed in their rate (2 or 3 Hz; blocked design), duration (2 or 3 s), and intelligibility (intelligible 16-channel or unintelligible 1-channel noise-vocoded speech). They were asked to press a button when they detected an irregularity in the stimulus rhythm (red targets). B. Performance (as d-prime) in the irregularity detection task, averaged across participants and shown for the main effects of intelligibility, duration, and rate. Error bars show standard error of mean (SEM), corrected for within-subject comparison (18). C. A rhythmic brain response measured during the presented sounds cannot distinguish true neural oscillations aligned to the stimulus from regular stimulus-evoked responses. However, only the oscillation-based model predicts a rhythmic response which outlasts the rhythmic stimulus. For each time point t throughout the trial, oscillatory phase was estimated based on a 1-s window centred on t (shaded grey). D. Inter-trial phase coherence (ITC) at time t is high when estimated phases are consistent across trials (left) and low otherwise (right). Note that the two examples shown differ in their 2-Hz ITC, but have similar induced power at the same frequency. E. ITC in the longer (3-s) condition, averaged across intelligibility conditions, gradiometers, and participants. Note that "time" (x-axis) refers to the centre of the 1-s windows used to estimate phase. ITC at 2 and 3 Hz, measured in response to 2 and 3 Hz sequences, were combined to form a rate-specific response index (RSR). The two time windows used for this analysis ("entrained" and "sustained") are shown in white (results are shown in Fig. 2). F. ITC as a function of neural frequency, separately for the two stimulation rates, and for the example time point shown as a black line in E.

Results

Experiment 1: Rhythmic intelligible speech produces sustained MEG oscillations

In Experiment 1, 21 participants listened to sequences of noise-vocoded (19) rhythmic speech (Fig. 1A), which were 2 or 3 seconds in duration and presented at one of two different rates (2 Hz and 3 Hz). Speech sequences consisted of 4, 6 or 9 one-syllable words, depending on sequence duration and speech rate. These words were either clearly intelligible or completely unintelligible and noise-like, depending on the number of spectral channels used during vocoding (16 or 1; see Materials and Methods).

In a subset of trials (12.5 %), one of the words in the sequence (red in Fig. 1A) was shifted towards another (± 68 ms), and participants were given the task to detect this irregularity in the stimulus rhythm. Replicating previous work (7), performance in this task (quantified as d-prime; see Materials and Methods; Fig. 1B) was enhanced for intelligible as compared to unintelligible speech (main effect of intelligibility in 3-way repeated-measures ANOVA, $F(1, 20) = 31.30$, $p < 0.0001$). We also found that irregularities were easier to detect if the sequence was longer (main effect of duration, $F(1, 20) = 32.39$, $p < 0.0001$) and presented at a faster rate (main effect of rate, $F(20) = 26.76$, $p < 0.0001$; no significant interactions).

Using MEG and EEG, we measured brain responses during the presented sounds and, importantly, in a subsequent, silent interval of several seconds that continued until the start of the next sequence (Fig. 1A,C). Due to its higher signal-noise ratio, we focused our initial analyses on the MEG data. We used inter-trial phase coherence (ITC) to quantify oscillatory brain responses (Fig. 1D). ITC makes use of the fact that, for each of the two speech rates, the timing of the presented speech sequences (relative to the "perceptual centre" of individual words, vertical lines in Fig. 1C) was identical across trials (see Materials and Methods). ITC therefore has the advantage of directly testing the predicted temporal evolution of the recorded signal (i.e. its phase), whereas power-based measures are focused on its amplitude (20). Fig. 1E shows ITC, separately for the two stimulus rates, and averaged across MEG sensors and participants. For one example time point, Fig. 1F shows ITC as a function of neural frequency.

Our hypothesis states that ITC at a given neural frequency is higher when that frequency corresponds to the stimulation rate than when it does not. For example, we expect that ITC at 2 Hz during (and after) the presentation of 2-Hz sequences (I in Fig. 1E,F) is higher than ITC at 2 Hz during (and after) 3-Hz sequences (II in Fig. 1E,F). By comparing ITCs across the two stimulus rates (I vs II and III vs IV in Fig. 1E,F), we thus developed a precise measurement of whether brain responses follow the rate of the stimulus, which we term the rate-specific response index (RSR; see Materials and Methods and formula in Fig. 1F). An RSR larger than 0 indicates a brain response that is specific to the stimulus rate. We then defined two time windows of interest (white in Fig. 1E). The first time window ("entrained") covered

the period in which sound sequences were presented while but avoiding sequence onset and offset. This period allows us to measure entrained responses (i.e. neural responses synchronised with an ongoing stimulus). A large RSR in this time window reflects a brain response aligned to the stimulus rhythm (irrespective of whether a true oscillation is involved). The other time window ("sustained") covered the silent interval between sequences while avoiding sequence offset. A large RSR in this time window is evidence for a sustained oscillatory response and, consequently, for the involvement of endogenous neural oscillations in generating stimulus-aligned entrained responses.

In the entrained time window, when averaged across all conditions, the RSR was clearly larger than 0 ($p < 0.001$; cluster-based correction), showing a typical auditory scalp topography (Fig. 2A). Using an ANOVA, we contrasted the RSR across conditions (Fig. 2B). We found a main effect of intelligibility ($p < 0.001$; cluster-based correction), revealing stronger rate-specific responses to intelligible speech in a cluster of left frontal sensors. There was no significant cluster for the main effect of duration, nor for the interaction intelligibility x duration.

Although the RSR was larger for intelligible speech, it was significantly larger than 0 (indicating the presence of an entrained response) for both intelligible ($p < 0.001$; cluster-based correction) and unintelligible speech ($p < 0.001$; cluster-based correction). The effect was localized to superior temporal regions (Fig. 2C) and was driven by rate-specific responses at both of the stimulated rates tested (Fig. 2D).

In the sustained time window, when averaged across all conditions, the RSR was significantly larger than 0 ($p = 0.05$; cluster-based correction) and maximal at left-lateralized parietal sensors (Fig. 2E). When contrasting RSR across conditions (Fig. 2F), we again found a main effect of intelligibility ($p = 0.01$; cluster-based correction), revealing stronger sustained rate-specific responses for intelligible speech. Importantly, these sustained responses were only significant (i.e. RSR > 0) after intelligible speech ($p = 0.01$; cluster-based correction); no significant cluster was found after unintelligible speech. Sustained effects after intelligible speech were localized to fronto-parietal brain regions, with a peak in left parietal regions (Fig. 2G).

To ensure that sustained oscillatory activity was not a result of aperiodic ("1/f") activity (21), which might differ between the two stimulus rates, we subtracted the "1/f component" from ITC measures of the sustained response (cf. (22)) by applying linear regression with reciprocal frequency (1/f) as a predictor of neural responses (dashed lines in Fig. 2H, left). We did this separately for the two stimulus rates, and re-computed the RSR using the residual (see Materials and Methods). This analysis confirms a sustained oscillatory response only after intelligible speech (Fig. S1), driven by rate-specific responses at both of the stimulated rates tested (Fig. 2H). Together, these effects demonstrate rhythmic brain

responses at a frequency corresponding to the rate of stimulation, which outlast the stimulation at parietal sensors, and are present after intelligible, but not unintelligible rhythmic speech.

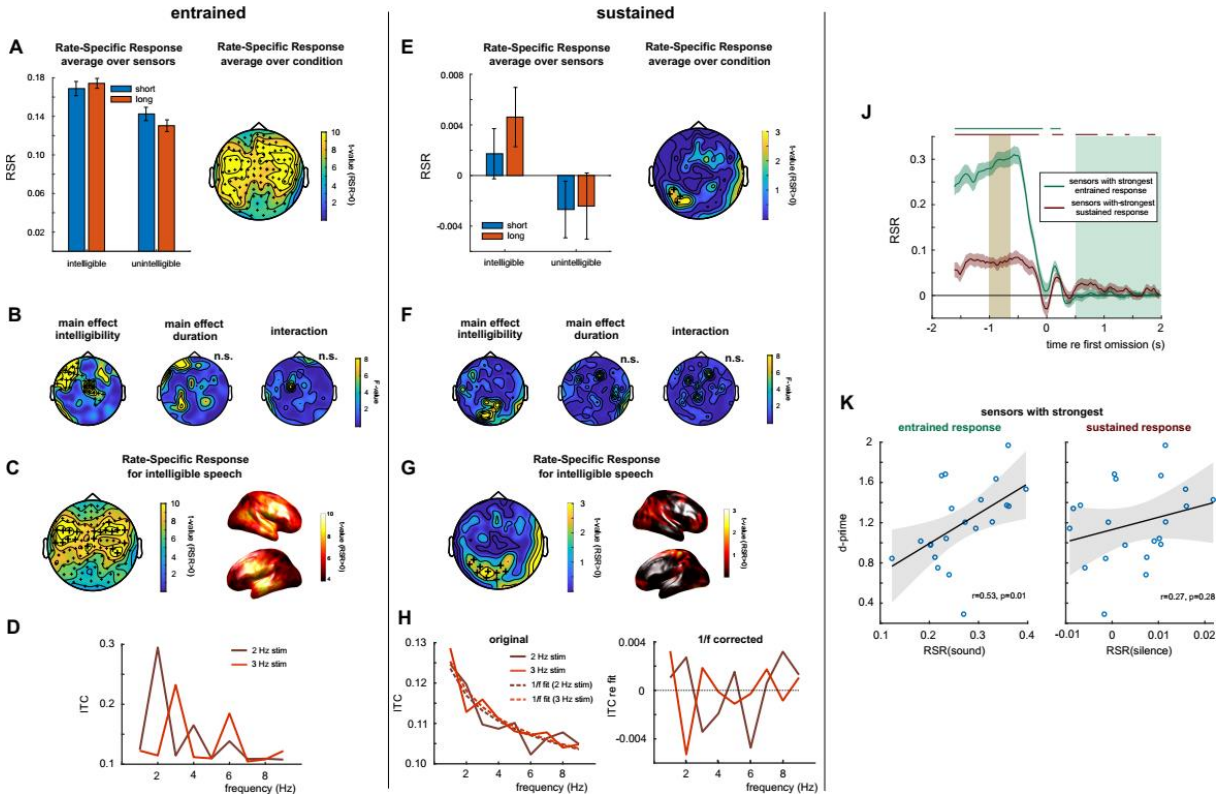

Figure 2. Results from Experiment 1. A-D. Results in the entrained time window (i.e. during rhythmic speech sequences; cf. Fig. 1E). Bars in panel A show RSR in the different conditions, averaged across gradiometers and participants. Error bars show SEM, corrected for within-subject comparison. The topography shows t-values for the comparison with 0, separately for the 102 gradiometer pairs, and after RSR was averaged across conditions. Topographies in B show F-values from an ANOVA, contrasting RSR across conditions. Topography and source plots in C show t-values for the comparison with 0 in the intelligible conditions. In all topographic plots, plus signs indicate the spatial extent of significant clusters from cluster-based permutation tests (see Materials and Methods). In A and C, this cluster includes all gradiometers (small plus signs). In C, larger plus signs show the 20 sensors with the highest RSR, selected for subsequent analyses. D shows ITC as a function of neural frequency, at these sensors and during intelligible speech, measured in response to 2 and 3 Hz sequences. Note that these ITC values were combined to form RSR shown in the other panels of Fig. 2, as detailed in Fig. 1F. E-H. Same as A-C, but for the sustained time window (i.e. after the rhythmic sequences; cf. Fig. 1E). For the right panel in H, a fitted "1/f" curve (shown as dashed lines in the left panel) has been subtracted from the data (see Materials and Methods). Note that, in panels D and H, the peaks correspond closely to the respective stimulus rates. J. RSR as a function of time, for the average of sensors marked with larger plus signs in C and G, respectively. Horizontal lines on top of the panel indicate an FDR-corrected p-value of <= 0.05 (t-test against 0) for the respective time point and sensor group. Shaded areas correspond to the two defined time windows (brown: entrained, green: sustained). Shaded areas around the curves show SEM. K. Correlation between RSR in the entrained (left) and sustained (right) time windows (for the same selected sensors as for G), respectively, and performance in the irregularity detection task (cf. Fig. 1B). Both RSR and performance were averaged across intelligibility and duration conditions; in addition, performance was averaged across rates. Shaded areas correspond to the confidence intervals of the regression lines.

To characterize the observed effects further, we selected the 20 sensors with the largest RSR during intelligible speech in the entrained time window (large plus signs in Fig. 2C; the significant cluster included all sensors). We also selected all sensors in the significant cluster obtained after intelligible speech in the sustained time window (Fig. 2G). Both selected sensor groups (based on entrained and sustained responses) showed a significant RSR throughout the entrained time window (horizontal lines in Fig. 2J; FDR-corrected). Importantly, the RSR at sensors selected to show a sustained response fluctuated at around the time of the first omitted word and then remained significantly above 0 for most of the sustained time window. Although the presence of a sustained RSR is expected (given the method used to select the sensors), this result gives us valuable insight into the timing of the observed effect. In particular, it excludes the possibility that the sustained effect is a short-lived consequence of the omission of an expected stimulus (see Discussion).

Finally, we correlated the RSR in both time windows (and at the selected sensors) with performance in the irregularity detection task (Fig. 2K). We found a significant correlation between RSR in the entrained time window and detection performance (Pearson's $r = 0.53$, $p = 0.01$), demonstrating behavioural relevance of entrained brain responses. Perhaps unsurprisingly, given that there is no temporal overlap between the sustained response and target presentation, individual differences in the sustained RSR did not show a significant correlation with individual differences in rhythm perception ($r = 0.27$, $p = 0.28$).

Experiment 2: tACS produces sustained rhythmic fluctuations in word report accuracy

In Experiment 1, we showed sustained oscillatory activity after rhythmic sequences of intelligible speech, indicating that endogenous neural oscillations are involved in generating speech-entrained brain responses. In Experiment 2, we tested whether tACS produces sustained rhythmic changes in speech perception; if observed this would not only provide an equivalent demonstration for tACS (i.e. that endogenous neural oscillations are entrained by transcranial electrical stimulation), but also show that these endogenous neural oscillations causally modulate perceptual outcomes.

Twenty participants were asked to report a single spoken, 16-channel vocoded target word, recorded rhythmically at 3 Hz, and embedded in background noise (Fig. 3A). The signal-noise ratio between target word and noise was adjusted for individual participants, ensuring similar task difficulty across participants and ensuring that effects of tACS were not obscured by floor or ceiling report accuracy (see Materials and Methods).

While participants performed this task, tACS was applied at 3 Hz over auditory regions, using the same configuration of bilateral circular and ring electrodes that yielded successful modulation of speech perception in (8) (see inset of Fig. 3A). In each trial, the target word was presented so that its "perceptual centre" (see Materials and Methods) falls at one of six different phase lags (red lines in Fig. 3A), relative to tACS. Prior to target presentation, tACS was applied for ~3, 4, or 5 seconds. Importantly, the target

word was presented either during tACS ("ongoing tACS"), which was turned off shortly afterwards, or immediately after tACS ("pre-target tACS"). We hypothesized that entrained neural activity due to tACS (irrespective of whether it involves endogenous oscillations; Fig. 3B) will produce a phasic modulation of speech perception in the ongoing tACS condition, as reported previously (8–10). However, in the pre-target tACS condition, such a phasic modulation can only be explained by sustained neural oscillations which lead to rhythmic changes in perception (Fig. 3C).

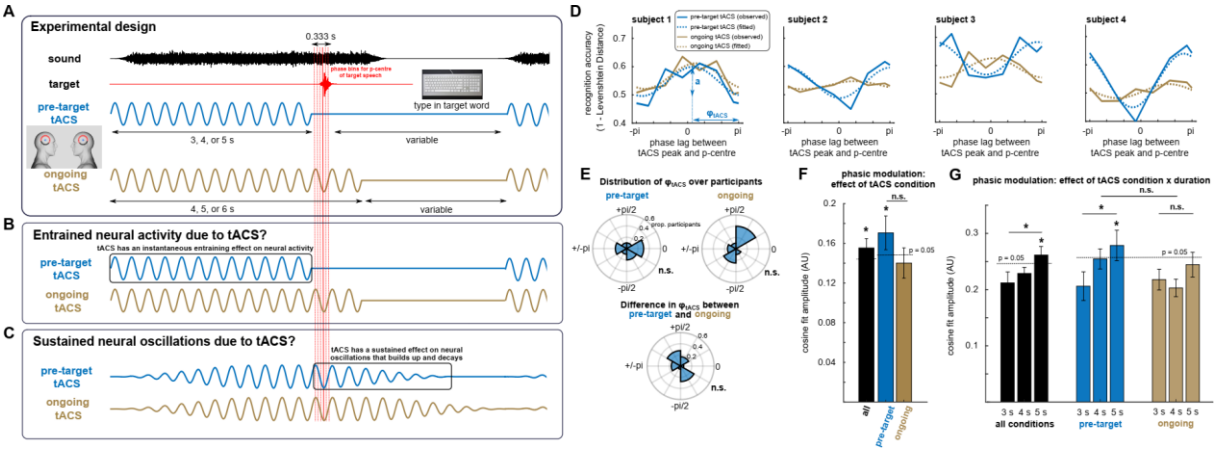

Figure 3. Experimental paradigm and main results from Experiment 2. A. Experimental paradigm. In each trial, a target word (red), embedded in noise (black), was presented so that its p-centre falls at one of six different phase lags (vertical red lines; the thicker red line corresponds to the p-centre of the example target), relative to preceding ("pre-target tACS") or ongoing tACS (which was then turned off). After each trial, participants were asked to type in the word they had heard. The inset shows the electrode configuration used for tACS in both conditions. B,C. Theoretical predictions. B. In the case of entrained neural activity due to tACS, this would closely follow the applied current and hence modulate perception of the target word only in the ongoing tACS condition. C. In the case that true oscillations are entrained by tACS, these would gradually decay after tACS offset and a "rhythmic entrainment echo" might therefore be apparent as a sustained oscillatory effect on perception even in the pre-target condition. D. Accuracy in the word report task as a function of phase lag (relative to tACS peak shown in A), averaged across tACS durations, and for four example participants. Phasic modulation of word report was quantified by fitting a cosine function to data from individual participants (dashed lines). The amplitude (a) of this cosine reflects the magnitude of the hypothesized phasic modulation. The phase of this cosine (φ_{tACS}) reflects the distance between its peak and the maximal phase lag of π. Note that the phase lag with highest accuracy for the individual participants, estimated based on the cosine fit, therefore corresponds to π-φ_{tACS}. E. Distribution of φ_{tACS} in the two tACS conditions, and their difference. F,G. Amplitudes of the fitted cosines (cf. amplitude a in panel D), averaged across participants. In F, cosine functions were fitted to data averaged over tACS duration (cf. panel D). In G, cosine functions were fitted separately for the three durations. For the black bars, cosine amplitudes were averaged across the two tACS conditions. Dashed lines show the threshold for statistical significance (p<=0.05) for a phasic modulation of task accuracy, obtained from a surrogate distribution (see Materials and Methods). Error bars show SEM (corrected for within-subject comparisons in F).

Accuracy in reporting the target word was quantified using Levenshtein distance (similar to the proportion of phonemes reported correctly (23); see Materials and Methods). When averaged across phase lags, word report accuracy was slightly higher in the pre-target tACS condition (0.50 ± 0.09, mean ± std) than in the ongoing tACS condition (0.49 ± 0.09), but not significantly different (t(19) = 1.67, p

= 0.11; repeated-measures t-test). This result indicates that the two tACS conditions did not reliably differ in their generic (i.e. phase-independent) effects on speech perception.

For each participant, and separately for the two tACS conditions, we determined how task accuracy varies with tACS phase lag (Fig. 3D). We then fitted a cosine function to data from individual participants (dashed lines in Fig. 3D). The amplitude of the cosine reflects how strongly speech perception is modulated by tACS phase. The phase of the cosine, labeled φ_{tACS}, reflects the distance between the peak of the cosine and the maximal phase lag tested (defined as π; Fig. 3D). For example, a φ_{tACS} of π would indicate highest word report accuracy at a tACS phase lag of 0.

Previous studies have reported that "preferred" tACS phase (leading to highest accuracy) varies across participants (7–10). Indeed, in neither of the two conditions did we find evidence for a non-uniform distribution of φ_{tACS} (Fig. 3E) across participants (Rayleigh's test for non-uniformity; pre-target tACS: z(19) = 0.64, p = 0.53; ongoing tACS: z(19) = 0.71, p = 0.50). We also failed to reveal a non-uniform distribution of the individual phase differences between conditions ($\varphi_{tACS(ongoing)}$ - $\varphi_{tACS(pre-target)}$; z(19) = 0.24, p = 0.79), indicating that the perceptual outcome in the ongoing and pre-target tACS conditions might not rely on identical neural processes.

To statistically evaluate the hypothesized phasic modulation of word report accuracy, we compared the observed cosine amplitudes (Fig. 3F,G) with a surrogate distribution – an approach which has recently been shown to be highly sensitive to detect such a phasic effect (24). The surrogate distribution was obtained by repeatedly shuffling experimental variables assigned to individual trials and extracting cosine amplitudes for each of those permutations. Here, these variables can refer to tACS phase lags, conditions, or durations, depending on the comparison of interest (see Materials and Methods).

We first pooled data over tACS durations (3, 4, and 5 s) before extracting cosine amplitudes (Fig. 3F). When tACS conditions were combined (i.e. their cosine amplitudes averaged), we found a significant phasic modulation of word report accuracy (z(19) = 2.80, p = 0.003). When conditions were analyzed separately, we found a significant phasic modulation of word report accuracy in the pre-target tACS condition (z(19) = 2.96, p = 0.002). This effect was not statistically reliable in the ongoing tACS condition (z(19) = 0.98, p = 0.16). However, the difference in modulation strength between tACS conditions was not significantly different from that obtained in a surrogate distribution (z(19) = 1.37, p = 0.17), indicating that the two conditions did not reliably differ in their efficacy of modulating speech perception.

We next tested whether the phasic modulation of speech perception depends on tACS duration (Fig. 3G). When tACS conditions were combined, we found an increase in phasic modulation of word report accuracy from 3-s tACS to 5-s tACS that was significantly larger than that observed in a surrogate distribution (z(19) = 1.82, p = 0.03). After five seconds of tACS, the phasic modulation was significant (z(19) = 2.36, p = 0.01), while the modulation was not statistically reliable after three seconds of stimulation (z(19) = -0.52, p = 0.70). When tACS conditions were analyzed separately, a significant

effect of duration was observed in the pre-target tACS condition ($z(19) = 1.86$, $p = 0.03$), but not in the ongoing tACS condition ($z(19) = 0.69$, $p = 0.24$). After five seconds of tACS, the phasic modulation of word report accuracy was significant in the pre-target tACS condition ($z(19) = 2.15$, $p = 0.016$), but not in the ongoing tACS condition ($z(19) = 1.17$, $p = 0.12$). However, when effects of duration (3-s tACS vs 5-s tACS) were compared across tACS conditions, we did not find a reliable difference between the two ($z(19) = 0.90$, $p = 0.37$), indicating that there was no significant interaction between tACS condition and duration.

Together, we found rhythmic changes in speech perception which outlast the period of tACS, and which depend on the duration of the preceding stimulation. Both of these findings are in line with an endogenous neural oscillation being entrained by tACS.

Experiment 1 vs 2: Phase of speech-entrained EEG predicts tACS effects in single participants

In line with previous research (7–10), we found that participants differ in the tACS phase leading to more or less accurate perception, reflected by φ_{tACS} (Fig. 3E). Although adapting tACS protocols to individual participants has been suggested as a crucial step to increase effect sizes and advance the field (25–27), neural parameters that can predict these individual differences remain elusive. Here, we report an analysis of combined data from 18 participants who participated in both our experiments. Rather than the MEG data reported earlier, we analysed the concurrent EEG data collected during Experiment 1 and relate this to tACS effects observed in Experiment 2 in the same participants. We anticipated that both tACS and EEG would be similarly affected by distortions in current flow in the skull and other, non-neural tissues (28). We therefore tested whether we can use EEG data to predict individual differences in φ_{tACS}.

In line with the MEG results reported earlier, EEG data in Experiment 1 showed a highly reliable rate-specific response (RSR) in the entrained time window (Fig. 4A; $p < 0.001$; cluster-based correction). The RSR in the sustained time window was largest at fronto-parietal electrodes, similar to our reported findings in MEG. However, this sustained effect was not statistically reliable (i.e. no significant clusters were obtained), either due to the lower signal-to-noise ratio of EEG or due to dipole orientation being less optimal for EEG.

Although the RSR combines ITC measured during two different stimulus rates (Fig. 1E,F), we here focused on EEG responses at 3 Hz in response to 3-Hz sequences, corresponding to the frequency of tACS in Experiment 2. Fig. 4B,C illustrates our analysis procedure for one example participant (Fig. 4B) and EEG electrode (Fig. 4B,C). For each EEG electrode, we extracted the phase of the 3-Hz response at each time point throughout the trial, and labeled it φ_{EEG} (Fig. 4B-II, green). We used Fast Fourier Transformation (FFT) to estimate φ_{EEG} (see Materials and Methods), which is equivalent to fitting a cosine at the frequency of interest (i.e. 3 Hz) to data in the analysis window (shaded grey in Fig.

4B-I) and extracting its phase. The value of φ_{EEG} therefore corresponds to the distance between each of the three peaks of the fitted cosine and the end of the corresponding cycle (defined as π; Fig. 4B-I)

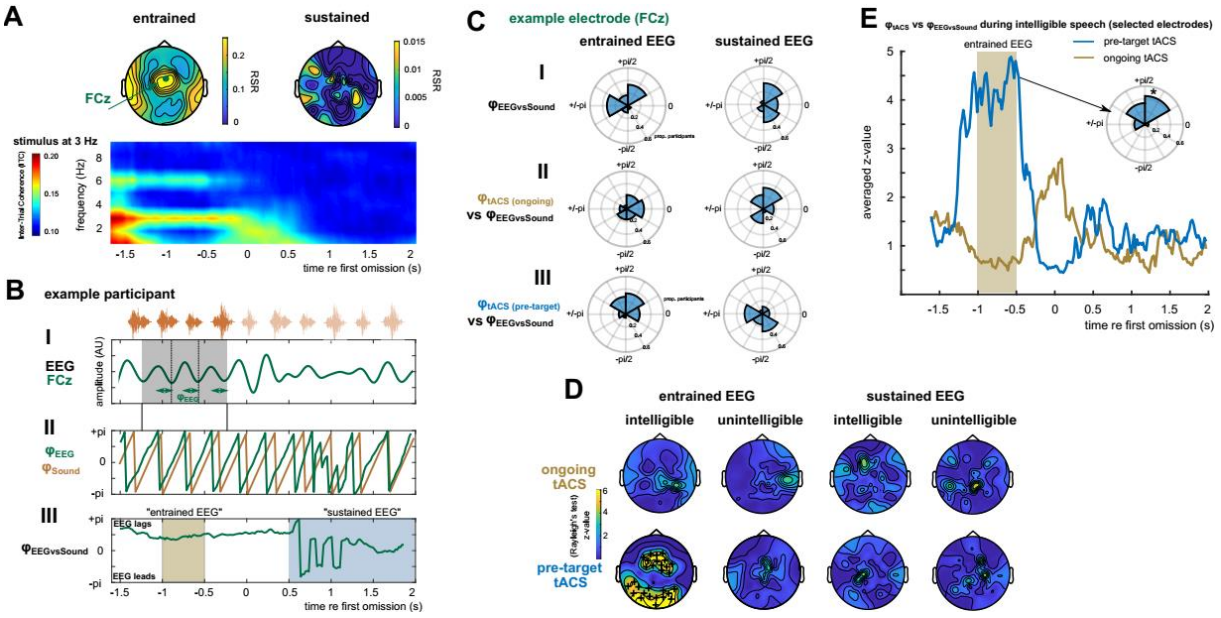

Figure 4. Combining Experiments 1 and 2. A. EEG results from Experiment 1. Topographies show RSR in the intelligible conditions. The green dot and arrow show the position of the electrode (FCz) with the highest RSR in the entrained time window. The time-frequency representation depicts ITC during 3-Hz sequences, averaged across EEG electrodes, participants, and conditions (cf. Fig. 1C). B. Illustration of methodological approach, using example data from one participant and electrode (FCz). B-I. Band-pass filtered (2-4 Hz) version of the EEG signal that has been used to estimate φ_{EEG} in the panel below (B-II). In practice, EEG phase at 3 Hz was estimated using FFT applied to unfiltered EEG data. Consequently, φ_{EEG} reflects the distance between the peaks of a cosine, fitted to data within the analysis window (shaded grey), and the end of each 3-Hz cycle (green arrows). B-II. φ_{EEG} (green; in the intelligible conditions and averaged across durations) and phase of the 3-Hz sequence (φ_{Sound}, orange). The latter is defined so that the perceptual centre of each word corresponds to phase π (see example sound sequence, and its theoretical continuation, on top of panel B-I). B-III. Circular difference between φ_{EEG} (green in B-II) and φ_{Sound} (orange in B-II), yielding $\varphi_{EEGvsSound}$. Given that φ is defined based on a cosine, a positive difference means that EEG lags sound. C. Distribution of individual $\varphi_{EEGvsSound}$, and its relation to φ_{tACS}. Data from one example electrode (FCz) is used to illustrate the procedure; main results and statistical outcomes are shown in panel D. C-I. Distribution of $\varphi_{EEGvsSound}$ (cf. B-III), extracted in the intelligible conditions, and averaged across durations and within the respective time windows (shaded brown and blue in B-III, respectively). C-II,III: Distribution of the circular difference between φ_{tACS} (Fig. 3E) and $\varphi_{EEGvsSound}$ (C-I). Note that a non-uniform distribution (tested in panel D) indicates a consistent lag between individual φ_{tACS} and $\varphi_{EEGvsSound}$. D. Z-values (obtained by means of a Rayleigh's test; see Materials and Methods), quantifying non-uniformity of the distributions shown in C-II,III for different combinations of experimental conditions. Plus signs show electrodes selected for follow-up analyses (FDR-corrected p <= 0.05). E. Z-values shown in D for intelligible conditions as a function of time, averaged across selected EEG sensors (plus signs in D). For the electrode with the highest predictive value for tACS (F3), the inset shows the distribution of the circular difference between φ_{tACS} and $\varphi_{EEGvsSound}$ in the pre-target condition, averaged within the entrained time window (shaded brown).

For each participant and EEG electrode, we determined how φ_{EEG} relates to the timing of the presented sound sequences (φ_{Sound}; Fig. 4B-II, blue). Assuming rhythmic EEG responses reliably following the presented sequences, the phase relation between EEG and sound (i.e. their circular difference) should

be approximately constant over time. This phase relation, labeled $\varphi_{EEGvsSound}$ (Fig. 4B-III), was therefore averaged within each of the two time windows of interest (entrained and sustained). The distribution of $\varphi_{EEGvsSound}$ across participants in these time windows is shown in Fig. 4C-I for the selected EEG electrode.

For each participant, EEG electrode, and the two time windows, we then calculated the (circular) difference between $\varphi_{EEGvsSound}$ and φ_{tACS} in the ongoing (Fig. 4C-II) and pre-target tACS conditions (Fig. 4C-III), respectively. Importantly, a non-uniform distribution would indicate a consistent lag between φ_{tACS} and $\varphi_{EEGvsSound}$ across participants. Fig. 4D shows the degree of non-uniformity of these distributions (as the z-values obtained in Rayleigh's test for non-uniformity; see Materials and Methods), for all EEG electrodes, and different combinations of conditions in the two experiments. We found that the phase relation between EEG and intelligible speech in the entrained time window significantly predicts φ_{tACS} in the pre-target tACS condition. This effect was maximal at fronto-central EEG electrodes (e.g., F3: $z(17) = 8.88$, $p = 0.003$, FDR-corrected for 70 electrodes). $\varphi_{EEGvsSound}$ was most predictive for φ_{tACS} around the presentation of the last word in the sequence (Fig. 4E). At the sensor with the strongest effect (F3), we observed a shift of ~90 degrees (corresponding to ~83.3 ms) between φ_{tACS} and $\varphi_{EEGvsSound}$ (inset in Fig. 4E).

This finding has important implications for future studies: Given the previous reports of tACS-induced changes in speech processing (7–11), tACS may be a promising tool to treat conditions associated with deficits in speech comprehension. However, individual differences in φ_{tACS} have so far hampered this goal – existing data suggest that different tACS phases will lead to optimal perception for each individual participant and extensive testing might therefore be needed to determine this optimal phase before further interventions. Based on the consistent phase shift between $\varphi_{EEGvsSound}$ and φ_{tACS} shown in Figure 4E, however, it should be possible to predict optimal tACS phase for single participants from EEG responses aligned to rhythmic intelligible speech. We tested this prediction in an additional analysis, as illustrated in Fig. 5 (see also Materials and Methods).

For this analysis, we selected EEG data from the entrained time window and the electrode (F3) with the strongest predictive value for pre-target tACS (Fig. 4D), and behavioural data from the same tACS condition. For each participant *i*, we determined their individual $\varphi_{EEGvsSound}$ (Fig. 5B) and used it to estimate their individual φ_{tACS} (Fig. 5C), based on the difference between the two that was observed on the group level (Fig. 5A,C). Importantly, for the latter, data from participant *i* was excluded, avoiding circularity of the procedure. For each participant, the estimated φ_{tACS} was then used to predict the tACS phase lag with highest accuracy in the word report task (blue dot in Fig. 5D,E). The behavioural data collected in Experiment 2 was re-aligned, relative to this predicted optimal phase lag (Fig. 5D; see Fig. S2 for individual re-aligned data from all participants). The outcome, averaged across participants, is

shown in Fig. 5F (blue). As intended, word report accuracy was highest at the predicted optimal phase lag (0 in Fig. 5F), and significantly higher than in the opposite phase bin (+/-π in Fig. 5F), which should lead to worst performance ($t(17) = 4.49$, $p < 0.001$). This result confirms that optimal tACS phases for speech perception can be estimated, exclusively based on individual EEG data (if the average difference between φ_{tACS} and $\varphi_{EEGvsSound}$ is known).

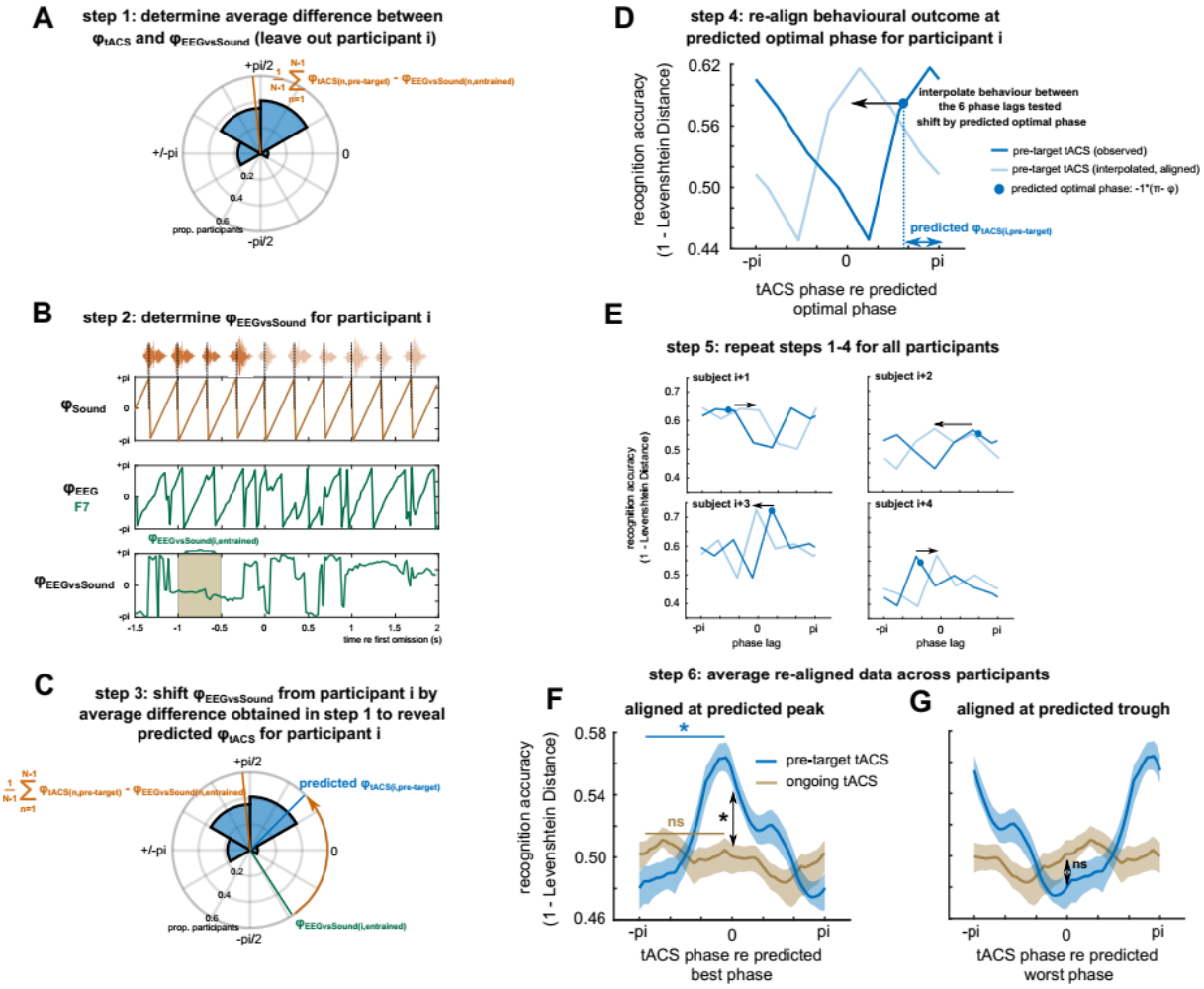

Figure 5. Predicted individual preferred tACS phases in the pre-target tACS condition from EEG data measured in the entrained time window at sensor F3. A, Step 1: For each participant *i*, data from all remaining participants was used to estimate the average difference between φ_{tACS} and $\varphi_{EEGvsSound}$. B, Step 2: $\varphi_{EEGvsSound}$ was determined for participant *i*. C, Step 3: This $\varphi_{EEGvsSound}$ was shifted by the phase difference obtained in step 1, yielding the predicted φ_{tACS} for participant *i*. D, Step 4: The predicted φ_{tACS} was used to estimate the tACS phase lag with highest perceptual accuracy for participant *i*, and the corresponding behavioural data was shifted so that highest accuracy was located at a centre phase bin. Prior to this step, the behavioural data measured at the six different phase lags was interpolated to enable re-alignment with higher precision. E, Step 5: This procedure was repeated for all participants. F, Step 6: The re-aligned data was averaged across participants (blue). For comparison, the procedure was repeated for the ongoing tACS condition (using EEG data from the same sensor). Given the non-significant results in Fig. 4D, the outcome reflects the null hypothesis of no enhanced accuracy at the predicted preferred tACS phase (brown). The coloured lines and asterisk/ns symbols reflect statistical comparisons between predicted best and worst phases within conditions, respectively. The black arrow and asterisk reflect the statistical comparison across conditions at the predicted best

phases. The shaded areas show SEM, corrected for within-subject comparison. G. Same as in F, but aligned at the predicted worst phase for word report accuracy.

Sustained oscillations produced by tACS enhance, but do not disrupt speech perception

It remains debated whether a phasic modulation of speech perception, produced by tACS, reflects a rhythmic enhancement or disruption of perception, or both (8–11, 29). Given that $\varphi_{EEGvsSound}$ was not predictive of φ_{tACS} in the ongoing tACS condition (Fig. 4D), we used data from the latter to test this question. We used the procedure illustrated in Fig. 5 (using data from the same EEG sensor F3) to predict optimal tACS phases in the ongoing tACS condition (see Materials and Methods). As $\varphi_{EEGvsSound}$ does not predict φ_{tACS} in this condition, any tACS-dependent modulation of task accuracy should be abolished by the re-alignment, and the re-aligned data (Fig. 5F, brown) should therefore reflect the null hypothesis, i.e. task outcome in the absence of a phasic modulation. Indeed, word report accuracy was not higher at the predicted optimal phase lag for the ongoing tACS condition than at the opposite phase lag ($t(17) = 0.08$, $p = 0.53$). We compared performance at the predicted optimal tACS phase between the two tACS conditions and found higher word report accuracy in the pre-target tACS condition ($t(17) = 3.48$, $p = 0.001$). Finally, for both conditions, we again re-aligned the behavioural data, but this time at the tACS predicted to be worst for performance (i.e. 180° away from the tACS phase predicted to be optimal for performance). Performance at the predicted worst tACS phase did not significantly differ between the two conditions ($t(17) = 1.34$, $p = 0.90$). These results show that the sustained phasic modulation of word report accuracy, produced by pre-target tACS, reflects a rhythmic enhancement of speech perception both relative to a non-optimal tACS phase and compared to EEG-aligned data from an ongoing tACS condition in which EEG data was not predictive of optimal tACS phase.

Discussion

In 1949, Walter & Walter (30) observed that rhythmic sensory stimulation produces rhythmic brain responses. Importantly, in their paper, when listing potential explanations for their observation, they distinguished "fusion of evoked responses giving an accidental appearance of rhythmicity" from "true augmentation or driving of local rhythms at the frequency of the stimulus". Now, more than 70 years later, it remains an ongoing debate whether "neural entrainment", brain responses aligned to rhythmic input, is due to the operation of endogenous neural oscillations or reflects a regular repetition of stimulus-evoked responses (16, 31–34). In two experiments, we provide clear evidence for entrained endogenous neural oscillations, by showing that rhythmic brain responses and rhythmic modulation of perceptual outcomes can outlast rhythmic sensory and electrical stimulation. We will discuss the implication of these sustained effects of sensory and electrical stimulation, before considering the functional interpretation of neural after-effects. We finish by discussing the potential for practical application of our combined EEG and tACS findings in supporting impaired speech perception.

Endogenous neural oscillations entrained by rhythmic sensory and electrical stimulation

Previous studies in a range of domains have similarly demonstrated sustained oscillatory effects after rhythmic sensory stimulation (summarized in (16)). Both perception and electrophysiological signals have been shown to briefly oscillate after a rhythmic sequence of simple visual (35–37) or auditory (38–40) stimuli, such as flashes or pure tones. A recent study showed that such a sustained rhythmic response occurs when preceded by a stimulus evoking the perception of a regular beat, but not when participants merely expect the occurrence of a rhythmic event (41). Although neural entrainment is widely explored in speech research (1, 2), we are only aware of one study reporting sustained oscillatory effects produced by human speech: Kösem et al (17) showed that, immediately after a change in speech rate, oscillatory MEG responses can still be measured at a frequency corresponding to the preceding speech (summarized in (15)). Our results in Experiment 1 are in line with this study and extend it by showing that (1) sustained oscillations produced by speech can be measured in silence and (2) are not observed for acoustically-matched speech stimuli that are unintelligible. Similar effects of intelligibility on neural entrainment have been described for combined tACS and fMRI: Neural responses in the STG to intelligible speech, but not to unintelligible speech, were modulated by tACS (7). In Experiment 1, we also replicated our previous MEG finding of more reliable stimulus-aligned responses to intelligible than unintelligible speech (5, 6). We further show that (1) rhythmic responses to intelligible speech persist after the offset of the speech stimulus and (2) this sustained effect is absent for acoustically-matched, unintelligible speech. Together, these MEG findings suggest that endogenous neural oscillations are active during neural entrainment, and that these oscillatory mechanisms are of particular importance for processing intelligible speech.

It is well established that the omission of an expected stimulus evokes a prominent neural response (42–45). One concern that could be raised regarding the present findings is whether our sustained effects could have been generated by an omission response rather than true oscillatory activity. Several aspects of our Experiment 1 suggest that omission-evoked responses are unlikely to explain the sustained effects of rhythmic stimulation: (1) omission responses would only lead to a sustained RSR if they were specific to the stimulation rate (i.e. if the omission leads to an increase in 2-Hz ITC after 2-Hz sequences and 3-Hz ITC after 3-Hz sequences); (2) sustained oscillatory activity after the end of a sequence lasts longer than would be expected from a single, punctate omission response (see Fig. 2J); (3) previous observations of omission responses show that these are largely generated in brain regions that were active while rhythmic stimuli were presented (44, 45), whereas our study showed sustained responses in other brain regions (compare scalp topographies and source distributions in Fig. 2C and 2G). These findings therefore suggest that sustained activity is generated by a true oscillatory response produced in response to intelligible speech.

Several studies have reported modulation of speech perception outcomes by tACS, and conclude that changes in neural entrainment, produced by varying the phase relation between tACS and speech

rhythm, are responsible (8–11). However, thus far these effects could reflect the rhythmic nature of the applied current, which might interfere with processing of speech presented with the same rhythm without any involvement of neural oscillations (15). In Experiment 2, we found sustained rhythmic fluctuations in speech perception that continued after the offset of tACS. Our results are an important extension of previous work as it suggests that: (1) modulation of speech perception can be due to the operation of neural oscillations entrained by tACS, and (2) sustained oscillatory effects after tACS can be measured in word report outcomes, and hence are causally relevant for speech perception. These findings for speech have precedent in other sensory modalities and brain regions. For example, a recent study (46) used tACS at 7 Hz to stimulate parietal-occipital regions and reported sustained rhythmic EEG responses at the frequency of electric stimulation. Although the functional role of these sustained neural effects for perceptual outcomes (in this case, perceptual integration) remain unclear, this study provides evidence for neural oscillations entrained by tACS that parallels the present work. The tACS method used here, in which perceptual effects are observed subsequent to the end of electrical stimulation are clearly amenable to further exploration in studies combining tACS and EEG.

Another line of evidence for endogenous oscillations entrained by a rhythmic stimulus comes from studies testing how brain responses vary as a function of stimulus rate and intensity (summarized in (16)). It is a clear prediction from classical physical models that the intensity required to entrain endogenous oscillations decreases when the rate of the entraining stimulus approaches their natural frequency (47–50). Indeed, this phenomenon, termed "Arnold Tongue", has recently been observed for visual stimulation (51). There is tentative evidence that tACS-induced responses behave in a similar way (summarized in (49)), but more studies are needed to substantiate this claim. Based on similar reasoning, entrainment effects should also be stronger when the system has "more time" to align with the external oscillator (49, 52). Our finding that tACS effects on perception increase with stimulation duration (Fig. 3G) is therefore clearly in line with oscillatory models. Importantly, such a behaviour was apparent in the pre-target tACS condition, in which effects of endogenous oscillations could be distinguished from those of other, potentially interfering neural processes. This result not only adds to existing demonstrations of endogenous oscillations entrained by tACS, it is also provides further evidence that point to entrained neural oscillations being more than just a passive response to rhythmic input. This idea is discussed in detail in the next section.

Rhythmic entrainment echoes – active predictions or passive after-effect?

In both our MEG and tACS experiments, we demonstrate that entrained neural and perceptual processes are more than a simple reflection of rhythmic input driving an otherwise silent system (Fig. 6A): Based on the observation of sustained oscillatory responses after the offset of stimulation, we conclude that an endogenous oscillatory system is involved in such entrained brain responses. What is the neural mechanism and functional role played by these rhythmic echoes of previously entrained responses

(hereafter, "entrainment echoes", cf. (46))? We here illustrate two different, but not mutually exclusive, models which can explain the observed entrainment echoes.

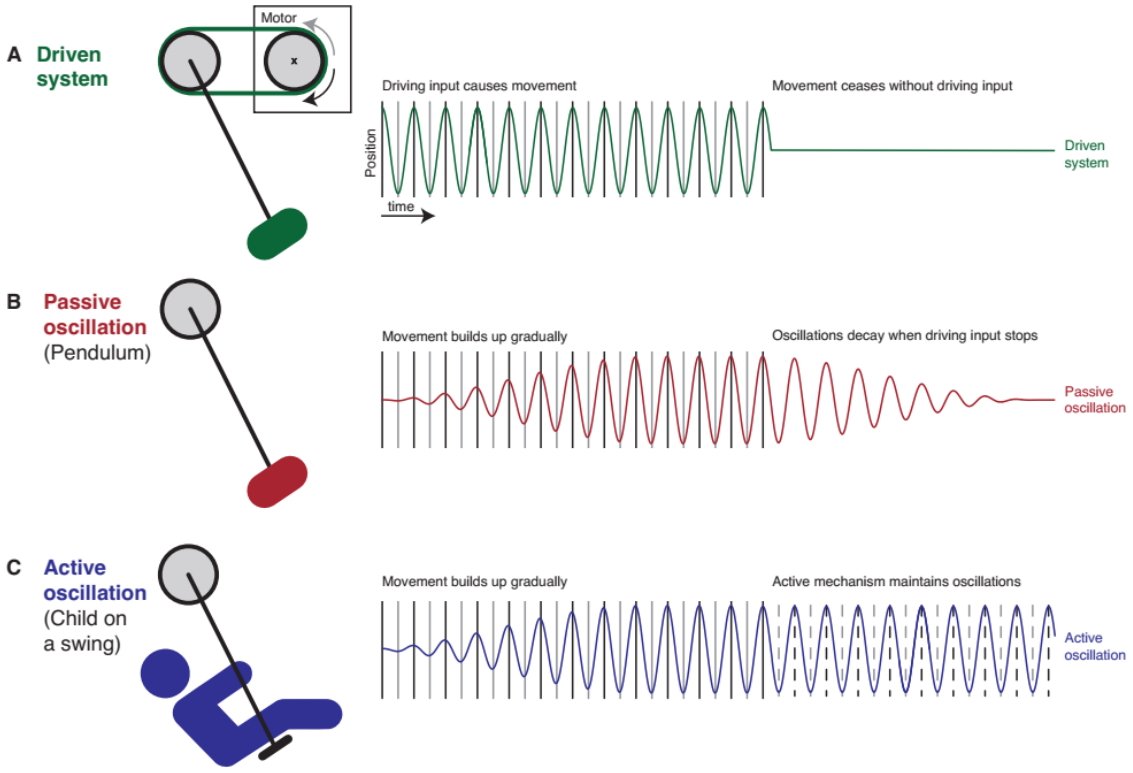

Figure 6. Three physical models that could be invoked to explain neural entrainment, and their potential to explain rhythmic entrainment echoes. A. In a system without any endogenous processes (e.g., neural oscillations), driving input would produce activity which ceases immediately when this input stops. B. A more direct account of rhythmic entrainment echoes is that endogenous neural oscillations resemble the operation of a pendulum which will start swinging passively when "pushed" by a rhythmic stimulus. When this stimulus stops, the oscillation will persist but decays over time, depending on certain "hard-wired" properties (similar to the frictional force and air resistance that slows the movement of a pendulum over time). C. Endogenous neural oscillations could include an active (e.g., predictive) component that controls a more passive process – similar to a child that can control the movement of a swing. This model predicts that oscillations are upheld after stimulus offset as long as the timing of important upcoming input (dashed lines) can be predicted. Note that, for the sake of clarity, we made extreme predictions to illustrate the different models. For instance, depending on the driving force of the rhythmic input, pendulum and swing could reach their maximum amplitude near-instantaneously in panels B and C, respectively, and therefore initially resemble the purely driven system shown in A. Similarly, it is possible that the predictive process (illustrated in C) operates less efficiently in the absence of driving input and therefore shows a decay similar to that shown by the more passive process (shown in B).

In one model, these rhythmic echoes reflect the passive reverberation of an endogenous neural oscillation that has previously been activated by a rhythmic stimulus. A physical analogy for this would be a pendulum that responds to a regular "push" by swinging back and forth, and that continues to produce a regular cyclical movement without external input until its kinetic energy has subsided (Fig. 6B). In the other model, stimulus-aligned oscillations are the result of an active mechanism that, through predictive processes, comes to align the optimal (high-excitability) oscillatory phase to the expected

timing of important sensory or neural events (12, 13). In this view, oscillatory activity can be actively maintained after stimulus offset and can persist for as long as these predictions are required. It is plausible that this active component is imposed onto a more "hard-wired", passive mechanism, that is oscillations might be entrained passively, but that this mechanism is under top-down control and can be adjusted if necessary. A physical analogy for this is the way in which a child will move on a swing if pushed, but can also control whether or not the movement of the swing is sustained after their helper stops pushing (Fig. 6C). The active mechanism, in this case, is the timing and amplitude of small movements that a sufficiently skilled child can coordinate with the movement of the swing to maintain oscillations without external help.

Several of our observations do point to an "active" component involved in generating rhythmic entrainment echoes, however, providing a definitive answer to this question remains for future studies. In both experiments, we found that the neural systems involved in producing sustained effects are distinct from those that are most active during the presence of the rhythmic stimulus. In Experiment 1, sustained MEG oscillations were maximal at parietal sensors and had a clearly different scalp topography and source configuration from typical auditory responses. In Experiment 2, individual tACS phase lags leading to highest word report accuracy *after* tACS offset were unrelated to those measured *during* tACS. Together, these findings are important as they speak against purely "bottom-up" or stimulus-driven generators of sustained oscillatory responses that merely continue to reverberate for some time after stimulus offset. Instead, they suggest that a distinct oscillatory network seems to be involved that might be specialized in "tracking" and anticipating important upcoming sensory events – potentially by adjusting and modulating a more passive, sensory processing system that aligns to rhythmic speech stimuli.

This proposal that top-down predictions for the timing of up-coming stimuli are achieved using neural oscillations is also in line with previous studies suggesting that neural predictions are fundamental for how human speech is processed by the brain (23, 53–55). It is possible that predictive oscillatory mechanisms are particularly strong for intelligible speech, and therefore upheld for some time when the speech stops. In contrast, unintelligible noise-like sequences, typically irrelevant in everyday situations, might lead to weaker predictions or shorter-duration sustained responses – explaining the results observed in Experiment 1.

Stronger rhythmic responses during intelligible than unintelligible speech (5, 6), as well as sustained oscillatory effects for speech sounds (17), have previously been shown in auditory brain areas. However, all of these studies measured neural effects during auditory input, which might bias localization of the neural responses towards auditory areas. Our study, in contrast, revealed sustained effects during post-stimulus silent periods at parietal sensors. This method might therefore yield a more precise estimate of

where these effects originate. Auditory input fluctuates rapidly, which requires the auditory system to quickly adapt its oscillations to changes in input (56, 57). Auditory input is represented more faithfully (i.e. less abstractly), and therefore on a faster time scale, in auditory brain regions than in "higher-level" ones (58). Thus, it is possible that oscillatory activity in the former involves more immediate responses, and hence disappears quickly after sound offset. In contrast, a more abstract representation of a rhythmic input – including phasic predictions about timing – might be more stable over time, and can remain present even after stimulus offset. This might be another reason to explain why our sustained oscillatory effects were found to be maximal at parietal sensors, potentially reflecting neural activity at a higher level of the cortical hierarchy.

Predicting tACS outcomes from EEG data – implications for future work and applications

It is a common observation that participants differ in how they respond to a given tACS protocol. For example, there is typically no consistent tACS phase which leads to highest perceptual accuracy for all participants (7–10). Individualizing brain stimulation protocols has therefore been proposed as a crucial step to advance the theoretical and practical application of this line of research (25–27). A recent study (59) reported that the phase relation between tACS and visual flicker modulates the magnitude of EEG responses to the flicker when tACS is turned off. Moreover, the individual "best" phase relation between tACS and flicker (leading to strongest EEG responses) was correlated with the individual phase relation between EEG and flicker. We replicate and extend this finding in a new modality by showing that the individual phase lag between EEG and intelligible speech can predict which tACS phase leads to more or less accurate perception in the same participant. Indeed, we found that EEG data from individual participants is sufficient to predict which tACS phase is optimal for perception, so long as the average lag between the two can be estimated even when using other, independent participants (Fig. 5). This result is important, as it shows that tACS can be adapted to individual brains based on EEG observations and establishes a method for aligning EEG and tACS findings for single participants. In an applied setting, these methods make the application of brain stimulation more efficient since the search for the most effective phase can be guided by EEG data rather than by trial and error. This finding therefore increases the potential for clinical or educational applications of tACS methods in future.

In previous work, using the same electrode configuration as applied in Experiment 2, we reported that tACS can only disrupt, and not enhance speech perception (8). We previously hypothesized that this is because tACS was applied simultaneously with rhythmic speech sequences, which themselves entrain brain activity. If neural entrainment to the speech sequences were already at the limit of what is physiologically possible, tACS might not be able to enhance it further. Importantly, in the current study, tACS was applied during non-rhythmic background noise, i.e. without any simultaneously entraining stimulus. Our finding of enhanced speech perception therefore supports the hypothesis that tACS can enhance neural entrainment if it is applied in the absence of a "competing" entraining stimulus. Together with the finding that tACS can be individualized, the protocol used here seems a promising method for

future technological applications in which tACS is used to enhance speech perception in a real-world setting.

In conclusion, we report evidence that endogenous neural oscillations are a critical component of brain responses that are aligned to intelligible speech sounds. This is a fundamental assumption in current models of speech processing (1) that we believe is only now clearly established by empirical evidence. We further show that tACS can modulate speech perception by entraining endogenous oscillatory activity. In this way we believe our work critically advances our understanding of how neural oscillations contribute to the processing of speech in the human brain.

Materials and Methods

Participants

24 participants were tested after giving informed consent in a procedure approved by the Cambridge Psychology Research Ethics Committee. 3 participants did not finish Experiment 1, leaving data from 21 participants (10 females; mean ± SD, 37 ± 16 years) for further analyses; 4 participants did not finish Experiment 2, leaving 20 participants for further analyses (11 females; 39 ± 15 years). 18 participants (9 females; 40 ± 15 years) finished both experiments.

All participants were native English speakers, had no history of hearing impairment, neurological disease, or any other exclusion criteria for MEG or tACS based on self-report.

Stimuli

Our stimuli consisted of a pool of ~650 monosyllabic words, spoken to a metronome beat at 1.6 Hz (inaudible to participants) by a male native speaker of British English (author MHD). These were time-compressed to 2 and 3 Hz, respectively, using the pitch-synchronous overlap and add (PSOLA) algorithm implemented in the Praat software package (version 6.12). This approach ensures that "perceptual centres", or "p-centres" (60) of the words were aligned to the metronome beat (see vertical lines in Fig. 1C) and, consequently, to rhythmic speech (in perceptual terms). Moreover, the well-defined rhythmicity of the stimulus allows a precise definition of the phase relation between stimulus and tACS (see below).

For Experiment 1 (Fig. 1A), these words were combined to form rhythmic sequences, which were 2 or 3 seconds long and presented at one of two different rates (2 or 3 Hz). Depending on the duration and rate of the sequence, these sequences therefore consisted of 4 (2 Hz / 2 s), 6 (3 Hz / 2 s and 2 Hz / 3s) or 9 words (3 Hz / 3s). Noise-vocoding (19) is a well-established method to produce degraded speech which varies in intelligibility, depending on the number of spectral channels used for vocoding. In Experiment 1, we used highly intelligible 16-channel vocoded speech and 1-channel noise-vocoded speech, which is a completely unintelligible, amplitude-modulated noise (for more details, see (7, 8)).

Importantly, noise-vocoding does not alter the rhythmic fluctuations in sound amplitude of the stimulus that are commonly assumed to be important for neural entrainment (61). Thus, acoustic differences in the broadband envelope between the two conditions cannot be responsible for differences in the observed neural responses.

For Experiment 2 (Fig. 3A), we presented participants with single 16-channel noise-vocoded target words, time-compressed to 3 Hz. These words were embedded in continuous noise with an average spectrum derived from all possible (~650) target words. The noise was presented for ~ 5-7 s. The target word occurred between 2 and 1.722 s before noise offset, depending on its phase lag relative to tACS (see Experimental Design and Fig. 3A). The noise was faded in and out at the beginning and end of each trial, respectively. All stimuli were presented to participants via headphones (through insert earphones connected via tubing to a pair of magnetically-shielded drivers in Experiment 1; ER-2 insert earphones in Experiment 2; Etymotic Research Inc., USA).

Experimental Design

In Experiment 1, while MEG/EEG data was recorded, participants listened to the rhythmic sequences (Fig. 1A) and pressed a button as soon as they detected an irregularity in the sequence rhythm (red in Fig. 1A). The irregularity was present in 12.5 % of the sequences and was produced by shifting one of the words (excluding first and last) in the sequence by ± 68 ms. Participants completed 10 experimental blocks of 64 trials each. For each block, the rate of the sequences was chosen pseudo-randomly and kept constant throughout the block. In each trial, the intelligibility (16- or 1-channel speech) and duration (2 or 3 s) of the sequence was chosen pseudo-randomly. Consequently, participants completed a total of 80 trials for each combination of conditions (rate x intelligibility x duration). Each of the sequences was followed by a silent interval in which sustained oscillatory responses were measured (Fig. 1C). These silent intervals were 2+x s long, where x corresponds to 1.5, 2, or 2.5 times the period of the sequence rate (i.e. 0.75, 1, or 1.25 s in 2-Hz blocks, and 0.5, 0.666, or 0.833 s in 3-Hz blocks). x was set to 2 in 50 % of the trials.

In Experiment 2, tACS was applied at 3 Hz and participants were asked to identify a target word embedded in noise, and report it after each trial using a standard computer keyboard. The start and end of each trial was signaled to participants as the fade in and out of the background noise, respectively (Fig. 3A). The next trial began when participants confirmed their response on the keyboard. We used an intermittent tACS protocol (cf. (59)), i.e. tACS was turned on and off in each trial. In two different tACS conditions, we tested how the timing of the target word relative to tACS modulates accuracy of reporting the target. In both conditions, the target word was presented so that its p-centre occurred at 3+y, 4+y, or 5+y seconds after tACS onset, chosen pseudo-randomly in each trial (red lines in Fig. 3A). y corresponds to one out of six tested phase delays between tACS and the perceptual center of the target word, covering one cycle of the 3-Hz tACS (corresponding to temporal delays between 66.67 ms and 344.45 ms, in steps of 55.56 ms). In the pre-target tACS condition, tACS was turned off y seconds before the

presentation of the target word. In the ongoing tACS condition, tACS remained on during the presentation of the target word and was turned off 1-*y* seconds after target presentation. In each trial, the background noise was faded in with a random delay relative to tACS onset (between 0 and 0.277 s). This ensured that the interval between noise onset and target was unrelated to the phase lag between tACS and target, avoiding potential alternative explanations for the hypothesized phasic modulation of word report by tACS. The background noise was faded out 1.5-y seconds after target presentation.

Participants completed 10 blocks of 36 trials each, leading to a total of 10 trials for each combination of conditions (tACS condition x duration x phase delay). Prior to the main experiment, they completed a short test in which the signal-noise ratio (SNR) between target word and background noise was adjusted and word report accuracy was assessed. During this test, no tACS was applied. Acoustic stimulation was identical to that in the main experiment, apart from the SNR, which was varied between -8 dB and 8 dB (in steps of 4 dB; 15 trials per SNR). From this pre-test, a single SNR condition at the steepest point on the psychometric curve (word report accuracy as a function of SNR) was selected and used throughout the main experiment (methods used for quantification of word report accuracy are described below in Statistical Analyses). This SNR was, on average -1.05 dB (SD: 1.75 dB).

For those participants who completed both experiments, Experiment 1 was always completed prior to Experiment 2, with, on average, 23 days between experiments (std: 30.88 days). However, all but two participants completed both experiments within one week of each other.

MEG/EEG Data Acquisition and Pre-processing (Experiment 1)

MEG was recorded in a magnetically and acoustically shielded room, using a VectorView system (Elekta Neuromag) with one magnetometer and two orthogonal planar gradiometers at each of 102 positions within a hemispheric array. EEG was recorded simultaneously using 70 Ag-AgCl sensors according to the extended 10–10 system and referenced to a sensor placed on the participant's nose. All data were digitally sampled at 1 kHz and band-pass filtered between 0.03 and 333 Hz (MEG) or between 0.1 and 333 Hz (EEG), respectively. Head position and electrooculography activity were monitored continuously using five head-position indicator (HPI) coils and two bipolar electrodes, respectively. A 3D digitizer (FASTRAK; Polhemus, Inc.) was used to record the positions of the EEG sensors, HPI coils, and ~70 additional points evenly distributed over the scalp relative to three anatomical fiducial points (the nasion and left and right preauricular points).

Data from MEG sensors (magnetometers and gradiometers) were processed using the temporal extension of Signal Source Separation (62) in MaxFilter software (Elekta Neuromag) to suppress noise sources, compensate for motion, and reconstruct any bad sensors.

MEG/EEG data were further processed using the FieldTrip software (63) implemented in MATLAB (The MathWorks, Inc.).

EEG data was high-pass filtered at 1 Hz and re-referenced to the sensor average. Noisy EEG sensors were identified by visual inspection and replaced by the average of neighbouring sensors. For MEG and EEG data separately, artefacts caused by eye movements, blinks, or heartbeat, were extracted using independent component analysis (ICA). ICA was applied to data down-sampled to 150 Hz. ICA components representing artefacts were identified visually and removed from the data at the original sampling rate of 1 kHz. The data were then epoched into trials from -3 s (longer condition) or -2 s (shorter condition) to +2.5 s, relative to the omission of the first word in each sequence (cf. Fig. 1C).

Electrical Stimulation (Experiment 2)

Current was administered using two battery-driven stimulators (DC-Stimulator MR, Neuroconn GmbH, Ilmenau, Germany). Each of the stimulators was driven remotely by the output of one channel of a high-quality sound card (Fireface UCX, RME, Germany); another output channel was used to transmit diotic auditory stimuli to the participants' headphones, assuring synchronization between applied current and presented stimuli.

We used a tACS electrode configuration that has produced a reliable modulation of word report in a previous study (8). This protocol entails bilateral stimulation over auditory areas using ring electrodes (see inset of Fig. 3A). Each pair of ring electrodes consisted of an inner, circular, electrode with a diameter of 20 mm and a thickness of 1 mm, and an outer, "doughnut-shaped", electrode with an outer and inner diameter of 100 and 75 mm, respectively, and a thickness of 2 mm. The inner electrodes were centered on T7 and T8 of the 10-10 system, respectively. The parts of the outer electrodes which overlapped with participants' ears were covered using electrically isolating tape. Electrodes were kept in place with adhesive, conductive ten20 paste (Weaver and Company, Aurora, CO, USA). Stimulation intensity was set to 1.4 mA (peak-to-peak) unless the participant reported stimulation to be unpleasant, in which case intensity was reduced (consequently, two participants were stimulated with 1.2 mA, one with 1.1 mA, and one with 1.0 mA). Current was not ramped up or down; we verified in preliminary tests that for sinusoidal stimulation this does not lead to increased current-induced sensations.

Sham stimulation was not applied in this experiment. Sensations produced by tACS are typically strongest at the onset of the electrical stimulation. Based on this notion, during sham stimulation, current is usually ramped up and down within several seconds, leading to similar sensations as during "true" tACS, but with no stimulation in the remainder of the trial or block (e.g., (64)). In the current experiment, we tested whether tACS applied for only several seconds leads to a phasic modulation of perception. Given the similarity of this approach to a typical sham stimulation condition, we did not expect that it would act as an appropriate control. Instead, we compared the observed tACS-induced modulation of speech perception with that obtained in a surrogate distribution, reflecting the null distribution (see Statistical Analyses).

Statistical Analyses

All analyses were implemented using custom MATLAB scripts and the toolbox for circular statistics (65), where appropriate.

Experiment 1

We first quantified rhythmic responses in our data using inter-trial phase coherence (ITC; Fig. 1D). At a given frequency and time, ITC measures the consistency of phase across trials (66, 67). ITC ranges between 0 (no phase consistency) and 1 (perfect phase consistency). Although some studies used spectral power to quantify oscillatory activity in rhythmic paradigms (e.g., (2)), ITC can be considered more appropriate in our case as it (1) as a measure based on phase, not power, directly takes into account the temporal structure of the data (20) and (2) is less affected by power differences across trials, which can bias results (e.g., trials with disproportionally high power can dominate the outcome). ITC at frequency *f* and time point *t* was calculated as follows:

$$ITC(f,t) = \left| \frac{1}{N} \sum_{n=1}^{N} e^{i(\varphi(f,t,n))} \right|$$

where $\varphi(f,t,n)$ is the phase in trial *n* at frequency *f* and time point *t*, and *N* is the number of trials.

φ was estimated using Fast Fourier Transform (FFT) in sliding time windows of 1 s (step size 20 ms; shown in grey in Fig. 1C,D), leading to a frequency resolution of 1 Hz. Note that, when the outcome of this time-frequency analysis is displayed (Figs. 1E, 2J, 4A,B,E, 5B), "time" always refers to the center of this time window.

ITC was calculated separately for each of the 204 orthogonal planar gradiometers and then averaged across the two gradiometers in each pair, yielding one ITC value for each of the 102 sensors positions. Data from magnetometers was only used for source localization (see below).

Our hypothesis states that we expect stronger rhythmic responses (i.e. ITC) at a given frequency when it corresponds to the rate of the (preceding) stimulus sequence (I and III in Fig. 1E,F) than when it does not (II and IV in Fig. 1E,F). We developed an index to quantify this rate-specificity of the measured brain responses (RSR). An RSR larger than 0 reflects a rhythmic response which follows the stimulation rate:

$$RSR_t = \big(ITC(f=2,r=2,t) - ITC(f=2,r=3,t)\big) + \\ (ITC(f=3,r=3,t) - ITC(f=3,r=2,t))$$

where *f* and *r* correspond to the frequency for which ITC was determined and sequence rate (both in Hz), respectively. For most analyses, *t* corresponds to a time interval within which ITC was averaged. Two such intervals were defined (white boxes in Fig. 1E): One to quantify rate-specific responses *during* the sequences, but avoiding sequence onset and offset (-1 to -0.5 s relative to the first omitted word), termed "entrained". The other to quantify rate-specific responses that *outlast* the sequences, and avoiding their offset (0.5 to 2 s relative to the first omitted word), termed "sustained".

To test whether rhythmic responses are present in these time windows and in the different conditions, we compared the RSR against 0, using Student's t-test. To test whether rhythmic responses are stronger in certain conditions, we subjected the RSR to a repeated-measures ANOVA with main factors intelligibility (16-channel vs 1-channel speech) and sequence duration (2 s vs 3 s).

These statistical tests were applied separately for each of the 102 MEG sensor positions (i.e. gradiometer pairs; Fig. 2A-H). Significant RSR (differences) were determined by means of cluster-based permutation tests (5000 permutations) (68). Sensors with a p-value <= 0.05 were selected as cluster candidates. Clusters were considered significant if the probability of obtaining their cluster statistic (e.g., sum of t-values) in the permuted dataset was <= 5 %.

Electro- or neurophysiological data analyzed in the spectral domain (e.g., to calculate ITC) often include aperiodic, non-oscillatory components with a "1/f" shape (21, 22). Not only can these 1/f components bias the outcome of spectral analyses (21, 22), they are also influenced by stimulus input (69). Consequently, it is possible that these aperiodic components differ between stimulus rates and therefore affect our RSR. To rule out such an effect, we repeated our RSR analysis, using ITC values corrected for 1/f components. For this purpose, a 1/f curve (22) was fitted to the ITC as a function of neural frequency, averaged within the time window of interest (dashed lines in Fig. 2H, left). This was done separately for each participant, sensor, stimulus rate, and experimental condition (intelligibility and duration), as these factors might influence the shape of the aperiodic component. Each of these fits was then subtracted from the corresponding data; the resulting residuals (Fig. 2H, right) reflect 1/f-corrected ITC values and were used to calculate RSR as described above. This procedure revealed prominent peaks at neural frequencies corresponding to the two stimulus rate (Fig. 2H, right), suggesting successful correction for aperiodic, non-oscillatory components. Given the absence of a pronounced 1/f component in the entrained time window (Fig. 2D), we here only show results for the sustained time window (Fig. 2H, Fig. S1).

Participants' sensitivity to detect an irregularity in the stimulus rhythm was quantified using d-prime (d'), computed as the standardized difference between hit probability and false alarm probability:

$$d' = z(p_{hit}) - z(p_{false\ alarm})$$

where, in a given condition, p_{hit} and $p_{false\ alarm}$ are the probability of correctly identifying an irregular sequence and falsely identifying a regular sequence as irregular, respectively.

To test whether performance in this task is correlated with rate-specific brain responses during or after the rhythmic sounds, we selected MEG sensors which responded strongly in the two time windows defined. In the entrained time window, all sensors were included in a significant cluster revealed by the analyses described above (Fig. 2C); we therefore selected the 20 sensors with the largest RSR. In the sustained time window, we selected all sensors which were part of a significant cluster (Fig. 2G). The RSR from those sensors (averaged within the respective time window) was correlated with performance

(d-prime), using Pearson's correlation. Even in conditions with relatively weak brain responses, these can still be related to task performance. For the correlation analysis, we therefore averaged both RSR and d-prime across conditions (intelligibility, duration, and rate, the latter for d-prime only).

RSR measured with MEG were source-localized using the following procedure. First, for each participant, MEG data was co-registered with their individual T1-weighted structural MRI, via realignment of the fiducial points. A structural MRI scan was not available for one participant, who was excluded from source analysis. Lead fields were constructed, based on individual MRI scans, using a single shell head model. Brain volumes were spatially normalized to a template MNI brain, and divided into grid points of 1 cm resolution. Source reconstruction was then performed, using a linear constrained minimum variance beamformer algorithm (LCMV (70)). Spatial filters were estimated, one for each of the two time windows of interest (entrained and sustained), and for each of the two neural frequencies that contribute to the RSR (2 Hz and 3 Hz). For each spatial filter, data from the two stimulus rates (2 Hz and 3 Hz) was combined, and single trials were band-pass filtered (2^{nd} order Butterworth) at the frequency for which the filter was constructed (2 Hz filter: 1-3 Hz; 3 Hz filter: 2-4 Hz). Data from gradiometers and magnetometers was combined. To take into account differences in signal strength between these sensor types, data from magnetometers was multiplied by a factor of 20 before the covariance matrix (necessary for LCMV beamforming) was extracted. Using other factors than 20 did not change results reported here. The spatial filters were then applied to fourier-transformed single-trial data at the frequency for which the filters were constructed (2 Hz and 3 Hz). The spatially filtered, fourier-transformed single-trials were then combined to form ITC, using the formula provided above. For *each* of the two stimulus rates (2 Hz and 3 Hz), this step yielded one ITC value per neural frequency of interest (2 Hz and 3 Hz), and for each of 2982 voxels inside the brain. These ITC values were then combined to RSR values, as described above.

Experiment 2

Participants' report of the target word was evaluated using Levenshtein distance (71), which is the minimum number of edits (deletions, insertions etc.) necessary to change a phonological representation of the participants responses into the phonology of the target word, divided by the number of phonemes in the word. Accuracy in the task was defined as 1 – Levenshtein distance; this measure varies between 0 and 1, where 1 reflects a perfectly reproduced target word (see (23) for details).

For each participant, tACS condition and duration separately, we tested how report accuracy varies with phase lag (corresponding to the delay between target word and tACS offset in the pre-target tACS condition, and to the actual tACS phase in the ongoing tACS condition; see Fig. 3A). This was done by fitting a cosine function to task accuracy as a function of phase lag (Fig. 3D), an approach which has recently been revealed as highly sensitive at detecting a phasic modulation of perception (24). The amplitude of the cosine (*a* in Fig. 3D) reflects how strongly performance varies as a function of phase lag. Note that *a* is always larger than 0. To test statistical significance, we therefore constructed a

surrogate distribution, which consists of amplitude values that would be observed in the absence of the hypothesized phase effect. For this purpose, phase lags were randomly assigned to trials and the analysis repeated to these shuffled datasets. This procedure was repeated 1000 times, yielding 1000 amplitude values for each experimental condition. The surrogate distribution was then compared with the single outcome obtained from the original, non-permuted data, resulting statistical (z-) values, according to:

$$z = (d-\mu) / \sigma$$

where d is the observed data, and μ and σ are mean and standard deviation of the surrogate distribution, respectively (24, 72). The phasic modulation of task accuracy, induced by tACS in a given condition, was considered reliable if the z-value exceeded a critical value (e.g., z = 1.645, corresponding to a significant threshold of $\alpha = 0.05$, one-tailed). We first tested for a phasic modulation of word report accuracy, irrespective of tACS duration (Fig. 3F). For this purpose, data was pooled over tACS duration before the cosine amplitudes were extracted. We then repeated the cosine fit procedure, separately for each duration (Fig. 3G). We analyzed the data separately for each tACS condition, as well as for their average. For the latter, cosine amplitude values were averaged since this does not require a consistent preferred phase for both conditions. For all statistical tests, values obtained from the surrogate distribution were treated in the same way as described for the original data.

To evaluate differences in phasic modulation of task accuracy between tACS conditions and durations, additional surrogate distributions were constructed by randomly assigning the variable of interest (i.e. tACS condition or tACS duration) to single trials and re-computing cosine amplitudes. To test for differences between tACS conditions, the difference in cosine amplitude between the two conditions was compared with the same difference in the surrogate distribution, using z-values as described above (two-tailed). Likewise, to test for differences between tACS durations, for each tACS condition separately and for their average, the difference in cosine amplitude between the longest (5-s) and shortest (3-s) durations was compared with the same difference in the surrogate distribution (one-tailed). To test for an interaction between tACS condition and duration, we first determined the difference in cosine amplitude between 5-s and 3-s tACS for each tACS condition, and then compared the difference between the two conditions with the same difference in the surrogate distribution (two-tailed).

Experiment 1 vs 2

Given the expected relationship between tACS and EEG (28), we tested whether the phase lag between tACS and target word, leading to particularly accurate or inaccurate responses in Experiment 2, can be predicted from the phase of EEG responses to rhythmic speech sequences in Experiment 1.

For this purpose, at each time point throughout the trial, EEG phase (φ_{EEG}, green in Fig. 4B-II) was extracted at 3 Hz (corresponding to the frequency at which tACS was applied in Experiment 2). Note that φ_{EEG} corresponds to $\varphi(f,t)$ defined above, where f = 3 Hz, and phase was averaged across trials

at time point *t*. As described above, φ was estimated using FFT and sliding analysis windows of 1 s. φ_{EEG} can therefore be understood as the phase of a 3-Hz cosine fitted to data within this 1-s window (shaded grey in Fig. 4B-I). The value of φ_{EEG} corresponds to the distance between each of the three cosine peaks and the end of the corresponding cycle (defined as π; arrow in Fig. 4B-I).

To obtain a more reliable estimate of phase, we combined phase estimates within each of the two time windows of interest (entrained and sustained). As averaging φ_{EEG} across time would lead to phase cancellation effects, we first determined, for each time point, the phase relation (i.e. circular difference) between EEG and the presented sequences. For the latter, φ_{Sound} (orange in Fig. 4B-II) was defined so that the perceptual centre of each word corresponds to π (compare example sounds on top of Fig. 4B-I with φ_{Sound} in Fig. 4B-II). Assuming a rhythmic EEG response that follows the presented sounds, the phase lag between φ_{EEG} and φ_{Sound} should be approximately constant across time. The circular difference between the two, labeled $\varphi_{EEGvsSound}$ (Fig. 4B-III) was therefore averaged within each of the two time windows. For the longer (3-s) sequences in Experiment 1, the entrained time window was extended to -2 to -0.5 s relative to the first omitted word (-1 to -0.5 s for shorter sequences).

For each the two tACS conditions, the phase of the cosine fitted to individual data, averaged across durations, was extracted (φ_{tACS} in Fig. 3D). φ_{tACS} reflects the position of the cosine peak (i.e. the "preferred" tACS phase, leading to highest accuracy), relative to the maximal phase lag tested (here: π).

For each participant, EEG electrode, and combination of conditions in the two experiments, we then extracted the circular difference between φ_{tACS} (Fig. 3D,E) and $\varphi_{EEGvsSound}$ (Fig. 4B-III,4C-I). The distribution of this difference (Fig. 4C-II,III) reveals whether there is a consistent phase lag between φ_{tACS} and $\varphi_{EEGvsSound}$ across participants. In this case, we would expect a non-uniform distribution, which was assessed with Rayleigh's test for non-uniformity (Fig. 4D). Despite potential differences in the *magnitude* of rhythmic brain responses, the different sequence durations tested in Experiment 1 should not differ in their *phase relation* to the sound. The $\varphi_{EEGvsSound}$ obtained in these conditions were therefore averaged. Finally, we selected 29 EEG sensors whose phase during intelligible speech was predictive (FDR-corrected p <= 0.05 in Rayleigh's test) for φ_{tACS} in the pre-target tACS condition (cf. Fig. 4D). The z-values, obtained from Rayleigh's test, were averaged and displayed as a function of time (i.e. not averaged within the two windows as described above).

We also used the obtained results to re-align behavioural outcomes in Experiment 2 relative to the predicted optimal tACS phase (leading to highest accuracy) in individual participants. We used a leave-one-participant-out procedure to avoid the inherent circularity in defining preferred phases or phase lags with the same data as used in the eventual analysis. This procedure is depicted in Fig. 5.

Step 1 (Fig. 5A): For each participant *i*, data from all remaining participants was used to estimate the average difference between φ_{tACS} (from the pre-target tACS condition) and $\varphi_{EEGvsSound}$. $\varphi_{EEGvsSound}$ was determined in the entrained time window, at electrode F3 (showing the highest predictive value for

φ_{tACS} in the pre-target condition). Step 2 (Fig. 5B): $\varphi_{EEGvsSound}$ was determined for participant *i*. Step 3 (Fig. 5C): The $\varphi_{EEGvsSound}$, obtained for participant *i* in step 2, was shifted by the average difference between φ_{tACS} and $\varphi_{EEGvsSound}$, obtained in step 1. This yielded the predicted φ_{tACS} for participant *i*. Step 4 (Fig. 5D): The predicted φ_{tACS} was used to estimate the tACS phase lag with highest perceptual accuracy for participant *i*. This phase lag was calculated as π-φ_{tACS}, based on the fact that φ_{tACS} reflects the distance between the peak of a fitted cosine and the maximal tACS phase lag (Fig. 3B). The behavioural data from participant *i* was then shifted by the predicted optimal phase lag, so that highest accuracy was located at a centre phase bin. As behavioural data was only available for six different phase lags, it was (linearly) interpolated between these data points (167 interpolated values between each phase lag) to enable a more accurate re-alignment of the data (note that the predicted φ_{tACS} depends on (1) the phase of the cosine fitted to individual data and (2) $\varphi_{EEGvsSound}$, neither of which are restricted to the six phase values tested).

Step 5 (Fig. 5E): Steps 1-4 were repeated, separately for each of the 18 participants. Step 6 (Fig. 5F). The re-aligned data was averaged across participants, with the hypothesis of highest accuracy at the predicted optimal phase lag for word report accuracy. This hypothesis was tested by comparing accuracy at this phase lag (0 in Fig. 5F) with accuracy at the one 180 ° (or π) away, using a one-tailed (given the clear one-directional hypothesis) paired t-test.

We also used this re-alignment procedure to test whether a rhythmic modulation of perception during or after tACS reflects enhancement or disruption of perception (or both). As our experimental protocol prevented the inclusion of the usual sham stimulation condition (see Electrical Stimulation), we based this analysis on the finding that φ_{tACS} was not reliably predicted by $\varphi_{EEGvsSound}$ in the ongoing tACS condition. We repeated the procedure described in the preceding paragraph; however, we used it to re-align behavioral outcome from the ongoing tACS condition to the phase lag predicted to be optimal for word report accuracy. Consequently, the only difference to the procedure described above is the use of φ_{tACS} obtained in the ongoing (not pre-target) tACS condition.

We compared accuracy at the predicted optimal tACS phase lag between the two tACS conditions. Given that $\varphi_{EEGvsSound}$is not predictive for φ_{tACS} in the ongoing tACS condition, any tACS-dependent changes in perception should be abolished by the re-alignment procedure, and the outcome reflects the null hypothesis. Consequently, higher accuracy at the predicted optimal phase lag in the pre-target tACS condition indicates an enhancement of speech perception, produced by tACS. This was tested by means of a one-tailed (given the clear one-directional hypothesis) paired t-test. Finally, we repeated the alignment procedure for both conditions, but this time aligned the behavioural data at the predicted worst phase lag for speech perception (i.e. 180° or π away from the predicted optimal phase). Again, we compared accuracy at this predicted worst phase lag between the two tACS conditions, using a one-tailed repeated-measures t-test. Lower accuracy at the predicted worst phase lag in the pre-target tACS condition indicates a disruption of speech perception, produced by tACS.

Acknowledgements

This work was supported by the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement number 743482, the British Academy/Leverhulme Trust (grant number SRG18R1\180733), and the Medical Research Council UK (grant number SUAG\008\RG91365). The authors thank Isobella Allard for support during pilot testing, Loes Beckers and Clare Cook for help with data acquisition, and Anne Kösem, Nina Suess and Nathan Weisz for advice on MEG source localization.

Declaration of Interests

The authors declare no competing interests.

References

1. A.-L. Giraud, D. Poeppel, Cortical oscillations and speech processing: emerging computational principles and operations. *Nat. Neurosci.* **15**, 511–517 (2012).

2. N. Ding, L. Melloni, H. Zhang, X. Tian, D. Poeppel, Cortical tracking of hierarchical linguistic structures in connected speech. *Nat. Neurosci.* **19**, 158–164 (2016).

3. J. E. Peelle, M. H. Davis, Neural Oscillations Carry Speech Rhythm through to Comprehension. *Front. Psychol.* **3**, 320 (2012).

4. B. Zoefel, R. VanRullen, The Role of High-Level Processes for Oscillatory Phase Entrainment to Speech Sound. *Front. Hum. Neurosci.* **9**, 651 (2015).

5. J. E. Peelle, J. Gross, M. H. Davis, Phase-locked responses to speech in human auditory cortex are enhanced during comprehension. *Cereb. Cortex* **23**, 1378–1387 (2013).

6. J. Gross, *et al.*, Speech rhythms and multiplexed oscillatory sensory coding in the human brain. *PLoS Biol.* **11**, e1001752 (2013).

7. B. Zoefel, A. Archer-Boyd, M. H. Davis, Phase Entrainment of Brain Oscillations Causally Modulates Neural Responses to Intelligible Speech. *Curr. Biol.* **28**, 401-408.e5 (2018).

8. B. Zoefel, I. Allard, M. Anil, M. H. Davis, Perception of Rhythmic Speech Is Modulated by Focal Bilateral Transcranial Alternating Current Stimulation. *J. Cogn. Neurosci.* **32**, 226–240 (2020).

9. L. Riecke, E. Formisano, B. Sorger, D. Başkent, E. Gaudrain, Neural Entrainment to Speech Modulates Speech Intelligibility. *Curr. Biol.* **28**, 161-169.e5 (2018).

10. A. Wilsch, T. Neuling, J. Obleser, C. S. Herrmann, Transcranial alternating current stimulation with speech envelopes modulates speech comprehension. *NeuroImage* **172**, 766–774 (2018).

11. M. Keshavarzi, M. Kegler, S. Kadir, T. Reichenbach, Transcranial alternating current stimulation in the theta band but not in the delta band modulates the comprehension of naturalistic speech in noise. *NeuroImage* **210**, 116557 (2020).

12. P. Lakatos, G. Karmos, A. D. Mehta, I. Ulbert, C. E. Schroeder, Entrainment of neuronal oscillations as a mechanism of attentional selection. *Science* **320**, 110–113 (2008).

13. C. E. Schroeder, P. Lakatos, Low-frequency neuronal oscillations as instruments of sensory selection. *Trends Neurosci.* **32**, 9–18 (2009).

14. J. Obleser, C. Kayser, Neural Entrainment and Attentional Selection in the Listening Brain. *Trends Cogn. Sci.* **23**, 913–926 (2019).

15. B. Zoefel, Speech Entrainment: Rhythmic Predictions Carried by Neural Oscillations. *Curr. Biol.* **28**, R1102–R1104 (2018).

16. B. Zoefel, S. ten Oever, A. T. Sack, The Involvement of Endogenous Neural Oscillations in the Processing of Rhythmic Input: More Than a Regular Repetition of Evoked Neural Responses. *Front. Neurosci.* **12** (2018).

17. A. Kösem, *et al.*, Neural Entrainment Determines the Words We Hear. *Curr. Biol.* **28**, 2867-2875.e3 (2018).

18. D. Cousineau, Confidence intervals in within-subject designs: A simpler solution to Loftus and Masson's method. *Tutor. Quant. Methods Psychol.*, 42–45 (2005).

19. R. V. Shannon, F. G. Zeng, V. Kamath, J. Wygonski, M. Ekelid, Speech recognition with primarily temporal cues. *Science* **270**, 303–304 (1995).

20. V. G. Rajendran, J. W. H. Schnupp, Frequency tagging cannot measure neural tracking of beat or meter. *Proc. Natl. Acad. Sci.* **116**, 2779–2780 (2019).

21. S. R. Cole, B. Voytek, Brain Oscillations and the Importance of Waveform Shape. *Trends Cogn. Sci.* **21**, 137–149 (2017).

22. M. Haller, *et al.*, Parameterizing neural power spectra. *bioRxiv*, 299859 (2018).

23. E. Sohoglu, M. H. Davis, Perceptual learning of degraded speech by minimizing prediction error. *Proc. Natl. Acad. Sci.* **113**, E1747–E1756 (2016).

24. B. Zoefel, M. H. Davis, G. Valente, L. Riecke, How to test for phasic modulation of neural and behavioural responses. *NeuroImage* **202**, 116175 (2019).

25. V. Romei, G. Thut, J. Silvanto, Information-Based Approaches of Noninvasive Transcranial Brain Stimulation. *Trends Neurosci.*, 782–795 (2016).

26. B. Zoefel, M. H. Davis, Transcranial electric stimulation for the investigation of speech perception and comprehension. *Lang. Cogn. Neurosci.* **32**, 910–923 (2017).

27. F. H. Kasten, K. Duecker, M. C. Maack, A. Meiser, C. S. Herrmann, Integrating electric field modeling and neuroimaging to explain inter-individual variability of tACS effects. *Nat. Commun.* **10**, 1–11 (2019).

28. S. Wagner, *et al.*, Using reciprocity for relating the simulation of transcranial current stimulation to the EEG forward problem. *NeuroImage* **140**, 163–173 (2016).

29. L. Riecke, B. Zoefel, Conveying Temporal Information to the Auditory System via Transcranial Current Stimulation. *Acta Acustica United with Acustica* **104**, 883-886 (2018).

30. V. J. Walter, W. G. Walter, The central effects of rhythmic sensory stimulation. *Electroencephalogr. Clin. Neurophysiol.* **1**, 57–86 (1949).

31. C. Keitel, C. Quigley, P. Ruhnau, Stimulus-driven brain oscillations in the alpha range: entrainment of intrinsic rhythms or frequency-following response? *J. Neurosci.* **34**, 10137–10140 (2014).

32. A. Capilla, P. Pazo-Alvarez, A. Darriba, P. Campo, J. Gross, Steady-state visual evoked potentials can be explained by temporal superposition of transient event-related responses. *PloS One* **6**, e14543 (2011).

33. S. Haegens, E. Zion Golumbic, Rhythmic facilitation of sensory processing: A critical review. *Neurosci. Biobehav. Rev.* **86**, 150–165 (2018).

34. K. B. Doelling, M. F. Assaneo, D. Bevilacqua, B. Pesaran, D. Poeppel, An oscillator model better predicts cortical entrainment to music. *Proc. Natl. Acad. Sci.*, 201816414 (2019).

35. E. Spaak, F. P. de Lange, O. Jensen, Local entrainment of α oscillations by visual stimuli causes cyclic modulation of perception. *J. Neurosci.* **34**, 3536–3544 (2014).

36. T. A. de Graaf, *et al.*, Alpha-band rhythms in visual task performance: phase-locking by rhythmic sensory stimulation. *PloS One* **8**, e60035 (2013).

37. K. E. Mathewson, *et al.*, Making waves in the stream of consciousness: entraining oscillations in EEG alpha and fluctuations in visual awareness with rhythmic visual stimulation. *J. Cogn. Neurosci.* **24**, 2321–2333 (2012).

38. P. Lakatos, *et al.*, The spectrotemporal filter mechanism of auditory selective attention. *Neuron* **77**, 750–761 (2013).

39. G. Hickok, H. Farahbod, K. Saberi, The Rhythm of Perception: Entrainment to Acoustic Rhythms Induces Subsequent Perceptual Oscillation. *Psychol. Sci.* **26**, 1006–1013 (2015).

40. F. C. Constantino, J. Z. Simon, Dynamic cortical representations of perceptual filling-in for missing acoustic rhythm. *Sci. Rep.* **7**, 1–10 (2017).

41. F. L. Bouwer, J. J. Fahrenfort, S. K. Millard, H. A. Slagter, A silent disco: Persistent entrainment of low-frequency neural oscillations underlies beat-based, but not memory-based temporal expectations. *bioRxiv*, 2020.01.08.899278 (2020).

42. H. C. Hughes, *et al.*, Responses of Human Auditory Association Cortex to the Omission of an Expected Acoustic Event. *NeuroImage* **13**, 1073–1089 (2001).

43. E. Sohoglu, M. Chait, Detecting and representing predictable structure during auditory scene analysis. *eLife* **5**, e19113 (2016).

44. T. Raij, J. P. Mäkelä, L. McEvoy, R. Hari, Human auditory cortex is activated by omissions of auditory stimuli. *Int. J. Psychophysiol.* **25**, 73 (1997).

45. I. SanMiguel, K. Saupe, E. Schröger, I know what is missing here: electrophysiological prediction error signals elicited by omissions of predicted ”what” but not ”when”. *Front. Hum. Neurosci.* **7** (2013).

46. R. Stonkus, V. Braun, J. R. Kerlin, G. Volberg, S. Hanslmayr, Probing the causal role of prestimulus interregional synchrony for perceptual integration via tACS. *Sci. Rep.* **6**, 1–13 (2016).

47. A. Pikovsky, *Synchronization: Universal Concept: A Universal Concept in Nonlinear Sciences*, 1st Paperback Ed edition (Cambridge University Press, 2008).

48. M. M. Ali, K. K. Sellers, F. Fröhlich, Transcranial alternating current stimulation modulates large-scale cortical network activity by network resonance. *J. Neurosci.* **33**, 11262–11275 (2013).

49. F. Fröhlich, Experiments and models of cortical oscillations as a target for noninvasive brain stimulation. *Prog. Brain Res.* **222**, 41–73 (2015).

50. J. Vosskuhl, D. Strüber, C. S. Herrmann, Non-invasive Brain Stimulation: A Paradigm Shift in Understanding Brain Oscillations. *Front. Hum. Neurosci.* **12** (2018).

51. A. Notbohm, J. Kurths, C. S. Herrmann, Modification of Brain Oscillations via Rhythmic Light Stimulation Provides Evidence for Entrainment but Not for Superposition of Event-Related Responses. *Front. Hum. Neurosci.* **10**, 10 (2016).

52. G. Thut, *et al.*, Rhythmic TMS causes local entrainment of natural oscillatory signatures. *Curr. Biol.* **21**, 1176–1185 (2011).

53. H. Blank, M. H. Davis, Prediction Errors but Not Sharpened Signals Simulate Multivoxel fMRI Patterns during Speech Perception. *PLoS Biol.* **14** (2016).

54. D. Norris, J. M. McQueen, A. Cutler, Prediction, Bayesian inference and feedback in speech recognition. *Lang. Cogn. Neurosci.* **31**, 4–18 (2016).

55. M. Davis, E. Sohoglu, "Three functions of prediction error for Bayesian inference in speech perception." in *Gazzaniga M., Mangun R., & Poeppel D. (Eds) The Cognitive Neurosciences, 6th Edition*, (MIT Press, 2020).

56. B. Zoefel, R. VanRullen, Oscillatory Mechanisms of Stimulus Processing and Selection in the Visual and Auditory Systems: State-of-the-Art, Speculations and Suggestions. *Front. Neurosci.* **11** (2017).

57. R. VanRullen, B. Zoefel, B. Ilhan, On the cyclic nature of perception in vision versus audition. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **369**, 20130214 (2014).

58. E. Edwards, E. F. Chang, Syllabic (~2-5 Hz) and fluctuation (~1-10 Hz) ranges in speech and auditory processing. *Hear. Res.* **305**, 113–134 (2013).

59. M. Fiene, *et al.*, Phase-specific manipulation of rhythmic brain activity by transcranial alternating current stimulation. *Brain Stimulat.* **13**, 1254–1262 (2020).

60. J. Morton, S. Marcus, C. Frankish, Perceptual Centers (P-centers). *Psychol. Rev.*, 405–408 (1976).

61. O. Ghitza, The theta-syllable: a unit of speech information defined by cortical function. *Front. Psychol.* **4**, 138 (2013).

62. S. Taulu, J. Simola, M. Kajola, Applications of the signal space separation method. *IEEE Trans. Signal Process.* **53**, 3359–3372 (2005).

63. R. Oostenveld, P. Fries, E. Maris, J.-M. Schoffelen, FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Comput. Intell. Neurosci.* **2011**, 156869 (2011).

64. F. H. Kasten, J. Dowsett, C. S. Herrmann, Sustained Aftereffect of α-tACS Lasts Up to 70 min after Stimulation. *Front. Hum. Neurosci.* **10** (2016).

65. , CircStat: A MATLAB Toolbox for Circular Statistics | Berens | Journal of Statistical Software (2017).

66. J. P. Lachaux, E. Rodriguez, J. Martinerie, F. J. Varela, Measuring phase synchrony in brain signals. *Hum. Brain Mapp.* **8**, 194–208 (1999).

67. S. Makeig, S. Debener, J. Onton, A. Delorme, Mining event-related brain dynamics. *Trends Cogn. Sci.* **8**, 204–210 (2004).

68. E. Maris, R. Oostenveld, Nonparametric statistical testing of EEG- and MEG-data. *J. Neurosci. Methods* **164**, 177–190 (2007).

69. B. J. He, Scale-free brain activity: past, present and future. *Trends Cogn. Sci.* **18**, 480–487 (2014).

70. B. D. Van Veen, W. van Drongelen, M. Yuchtman, A. Suzuki, Localization of brain electrical activity via linearly constrained minimum variance spatial filtering. *IEEE Trans. Biomed. Eng.* **44**, 867–880 (1997).

71. V. I. Levenshtein, Binary Codes Capable of Correcting Deletions, Insertions and Reversals. *SPhD* **10**, 707 (1966).

72. R. VanRullen, How to Evaluate Phase Differences between Trial Groups in Ongoing Electrophysiological Signals. *Front. Neurosci.* **10** (2016).

SUPPLEMENTAL INFORMATION

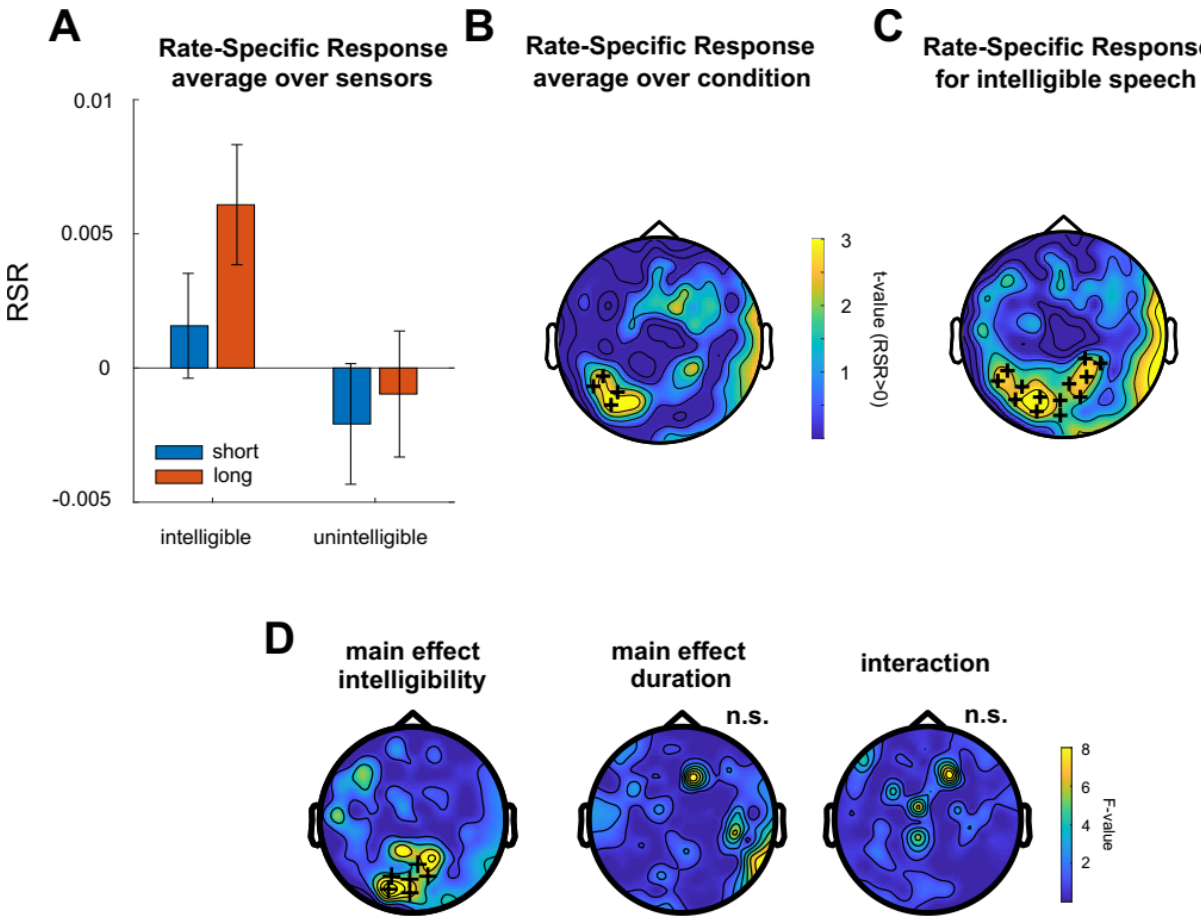

Figure S1. **Rate-specific responses (RSR) in sustained time window after correction for 1/f component**. Same as in Fig. 2E-G, but using 1/f-corrected Inter-Trial Coherence (shown in Fig. 2H) to calculate RSR. Same conventions as for Fig. 2.

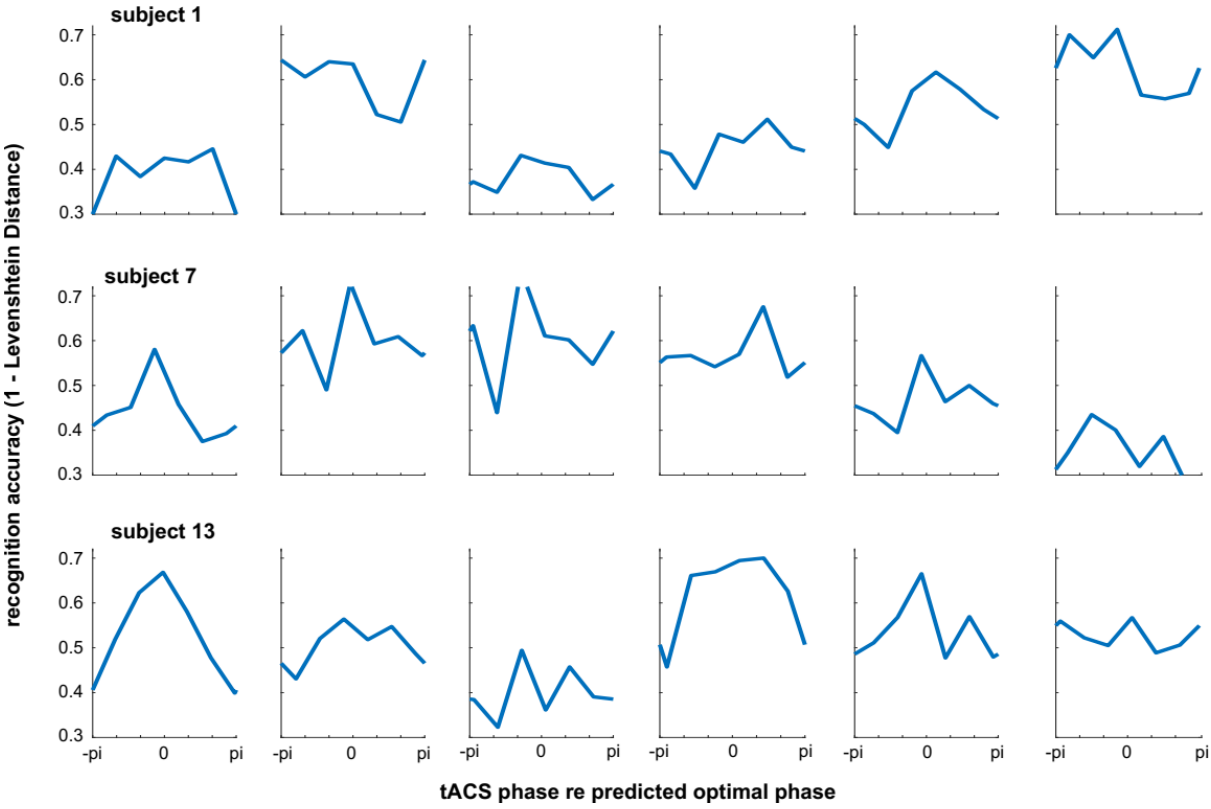

Figure S2. **Data from all individual participants, re-aligned to predicted optimal tACS phase.** Same as Fig. 5D,E, but for all 18 participants who were included in the analysis. Note that the average across participants is shown in Fig. 5F.