Neural Entrainment to Speech Modulates Speech Intelligibility

Lars Riecke, Elia Formisano, Bettina Sorger, Deniz Başkent, Etienne Gaudrain

Article

Current Biology

Neural Entrainment to Speech Modulates Speech Intelligibility

Highlights

- Transcranial stimulation with speech-shaped currents influences auditory processing
- Speech-brain entrainment modulates speech intelligibility
- Speech-brain entrainment and speech intelligibility interact reciprocally

Authors

Lars Riecke, Elia Formisano, Bettina Sorger, Deniz Başkent, Etienne Gaudrain

Correspondence

l.riecke@maastrichtuniversity.nl

In Brief

Riecke et al. study how humans can recognize speech. Using electric brain stimulation, they find that synchronization of ongoing brain activity with the rhythm of auditory speech modulates the intelligibility of this speech. This implies that the brain can employ its ongoing temporal activity as a critical instrument in speech recognition.

Riecke et al., 2018, Current Biology *28*, 1–9
January 22, 2018
https://doi.org/10.1016/j.cub.2017.11.033

CellPress

Please cite this article in press as: Riecke et al., Neural Entrainment to Speech Modulates Speech Intelligibility, Current Biology (2017), https://doi.org/10.1016/j.cub.2017.11.033

Current Biology
Article

Neural Entrainment to Speech Modulates Speech Intelligibility

Lars Riecke,[1,5,*] Elia Formisano,[1] Bettina Sorger,[1] Deniz Başkent,[2,4] and Etienne Gaudrain[2,3,4]
[1]Department of Cognitive Neuroscience, Faculty of Psychology and Neuroscience, Maastricht University, 6229 EV Maastricht, the Netherlands
[2]Department of Otorhinolaryngology/Head and Neck Surgery, University Medical Center Groningen, University of Groningen, 9700 RB Groningen, the Netherlands
[3]CNRS UMR 5292, Lyon Neuroscience Research Center, Auditory Cognition and Psychoacoustics, Inserm UMRS 1028, Université Claude Bernard Lyon 1, Université de Lyon, 69366 Lyon Cedex 07, France
[4]These authors contributed equally
[5]Lead Contact
*Correspondence: l.riecke@maastrichtuniversity.nl
https://doi.org/10.1016/j.cub.2017.11.033

SUMMARY

Speech is crucial for communication in everyday life. Speech-brain entrainment, the alignment of neural activity to the slow temporal fluctuations (envelope) of acoustic speech input, is a ubiquitous element of current theories of speech processing. Associations between speech-brain entrainment and acoustic speech signal, listening task, and speech intelligibility have been observed repeatedly. However, a methodological bottleneck has prevented so far clarifying whether speech-brain entrainment contributes functionally to (i.e., causes) speech intelligibility or is merely an epiphenomenon of it. To address this long-standing issue, we experimentally manipulated speech-brain entrainment without concomitant acoustic and task-related variations, using a brain stimulation approach that enables modulating listeners' neural activity with transcranial currents carrying speech-envelope information. Results from two experiments involving a cocktail-party-like scenario and a listening situation devoid of aural speech-amplitude envelope input reveal consistent effects on listeners' speech-recognition performance, demonstrating a causal role of speech-brain entrainment in speech intelligibility. Our findings imply that speech-brain entrainment is critical for auditory speech comprehension and suggest that transcranial stimulation with speech-envelope-shaped currents can be utilized to modulate speech comprehension in impaired listening conditions.

INTRODUCTION

In naturally produced auditory speech, intervals containing strong phonetic content (e.g., syllables) alternate quasi-rhythmically with intervals containing less phonetic content (e.g., silences). This phonetic rhythm is conveyed by the slow (~1–8 Hz) temporal fluctuations of the acoustic speech signal, called speech envelope [1, 2]. Speech-envelope information is critical for intelligibility (e.g., [3, 4]). It evokes a marked "envelope-following" neural response in the auditory cortex and thereby temporally aligns ongoing auditory cortical activity in the delta/theta (1–8 Hz) range to it. This phenomenon, called "speech-brain entrainment" (hereafter referred to as "speech entrainment" for brevity), has been observed reliably with various neuroelectromagnetic recording methods (invasive and non-invasive electroencephalography and magnetoencephalography), even at the single-trial level. Speech entrainment is evoked by prominent landmarks in the temporal envelope (e.g., [5]) and/or the linguistic structures [6–8] of an acoustic speech input. Moreover, it is modulated by selective attention and temporal expectancies (e.g., [9–14]) via endogenous cortical oscillatory activity, making it a powerful instrument for the brain to actively select linguistic information [15]. Whereas such bottom-up and top-down contributions to speech entrainment are being increasingly understood and incorporated in neurocognitive models of speech processing/perception [1, 16–21], the correlational nature of the applied study designs has hampered disentangling the putative functional roles of speech entrainment and intelligibility. Although covariations have been observed repeatedly (e.g., [5, 22, 23]), it could not be explicitly tested whether speech entrainment functionally contributes to (i.e., causes) intelligibility, as often presumed, or is merely an epiphenomenon of it.

To address this unresolved question, we tested in the present study the putative causal role of speech entrainment in speech intelligibility. We circumvented the previous methodological bottleneck with a novel methodological approach that we refer to as "speech-envelope-shaped transcranial current stimulation" ("envTCS"). EnvTCS involves the silent and non-invasive (scalp-based) application of an electric current carrying speech-envelope information. Because neural excitability in cortex follows the waveform of an externally applied current (e.g., [24, 25]), application of envTCS over auditory cortical regions involved in speech entrainment may bias bottom-up auditory speech processing toward the specific temporal pattern inherent in the applied speech-envelope-shaped current (e.g., [26]). In particular, the relative timing of the envTCS-following neural

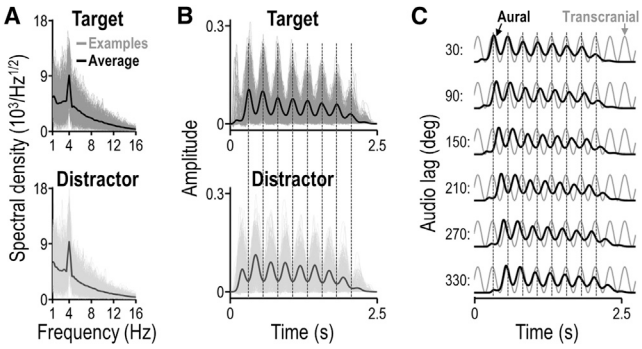

Figure 1. Stimulus Characteristics and Experimental Design for the Two-Talker Experiment

(A) Magnitude spectrum of the modified speech envelope, for target talker (top) and distracting talker (bottom). Thin lines represent individual sentences, and the thick line represents their average. The plots highlight the prominent 4-Hz rhythm of the aurally presented speech signals.

(B) Modified speech envelopes underlying the modulation spectra shown in (A). Envelopes of target (top) and distractor (bottom) sentences were anti-phasic; thus, portions containing strong phonetic content (e.g., syllables) alternated across talkers in the two-talker stimuli. See also Audio S1.

(C) Sketch of the experimental design. The six experimental conditions (rows) were characterized by the delay by which the aurally presented target-talker envelope (black waveform; same as B, top) lagged behind the transcranially applied 4-Hz alternating current (gray waveform). This experimental "audio-lag" manipulation served to induce 4-Hz variations in the strength of neural entrainment to target-speech-evoked responses. These variations were predicted to cause corresponding changes in speech intelligibility performance.

excitability and auditory speech-evoked neural responses should determine the strength of speech entrainment: when the envTCS is temporally aligned with these bottom-up responses, the latter should be enhanced and thereby speech entrainment should be strengthened, compared with when the two are misaligned.

We applied envTCS simultaneously with aurally presented conversational speech, and we varied their relative timing with the aim to experimentally manipulate the strength of speech entrainment. Given previously observed effects of aurally presented speech envelope on both neural entrainment and intelligibility (see above), we conjectured that our transcranial manipulation of entrainment alone (i.e., in the absence of acoustic and task-related changes) would suffice to induce systematic changes in intelligibility. The results from two speech-recognition experiments support this prediction in both a cocktail-party-like scenario and a listening situation devoid of aural speech-amplitude envelope input. These findings provide strong evidence that neural speech entrainment plays indeed a causal role in speech intelligibility.

RESULTS

Two-Talker Experiment

The first experiment was designed to identify whether neural speech entrainment modulates speech intelligibility in a two-talker situation. Speech materials were sentence recordings from a male and a female native Dutch talker with a speech rhythm that we artificially enhanced and fixed to a critical frequency f_c = 4 Hz (Figure 1A), which corresponds closely to the average syllable rate in Dutch. We mixed the speech signals in such a way that the two talkers' speech rhythms alternated; see Figure 1B and Audio S1. Participants focused on the male talker (target) in the mixture, while we simultaneously applied envTCS with a sinusoidal current alternating at the speech rate (equivalent to f_c transcranial alternating current stimulation [TACS]). Based on prior electric-field simulations and behavioral findings [27], we presumed this electric stimulation to entrain delta/theta neural excitability cycles in the human auditory cortex, a region that has been associated with speech entrainment [28, 29]. We systematically varied the relative timing of the alternating current and the aurally presented speech rhythm of the target talker in a cyclical manner (Figure 1C). We predicted that, if speech entrainment contributes functionally to speech intelligibility, this experimental "audio-lag" manipulation should induce cyclical f_c changes in the strength of entrainment to target-speech-evoked cortical responses, which would be observable as a corresponding f_c cycle in target-speech-recognition performance. Alternatively, if speech entrainment is merely an epiphenomenon of speech intelligibility, no systematic change in behavioral performance should be observed.

To test this, we assessed listeners' speech performance as a function of the strength of target-speech entrainment, assuming that envTCS entrained neural excitability as described above. Under our hypothesis that stronger speech-brain entrainment leads to more effective auditory speech processing, we associated maximum entrainment strength with the audio lag for which individual participants performed best ("best lag"; Figure 2A) and lower entrainment strengths with more distant lags while preserving the circular structure of the lags. We then tested whether the resulting performance waveform exhibited characteristics of the predicted f_c cycle, i.e., whether lags near the best lag (presumed to elicit strong entrainment) revealed an "excitatory" f_c half-cycle (i.e., relatively good performance) compared with more distant lags, which should reveal the opposite.

Speech Entrainment Modulates Speech Intelligibility in a Two-Talker Situation

Listeners correctly recognized on average 56.2% ± 2.1% (mean ± SEM) of the target talker's words. Conforming to our predictions, their performance varied significantly across the presumed entrainment strengths (main effect of best-lag distance: $F_{4,80}$ = 3.02, η^2 = 0.02, p = 0.02); see Figures 2B and 2C. To avoid circular reasoning, the trivial peak performance at the best lag was excluded from this analysis. Inspection of the performance waveforms revealed indeed better performance at all lags presumed to elicit strong entrainment (best-lag distances −60° and 60°) compared with all lags presumed to elicit weak entrainment (best-lag distances 120°–240°). Although, on average, the worst performance was not associated with the most distant lag (180°), suggesting contributions from neural oscillations beyond f_c [30], these observations matched well the characteristics of the predicted f_c cycle. We verified this notion by comparing the average performance during the presumed excitatory half-cycle (−60° and 60°) with that during the opposite half-cycle (120° and 240°), which revealed a significant difference of on average 3.0 ± 0.8 percentage points in the predicted direction (t_{20} = 3.94, d = 0.86, corrected p = 0.0012); see Figure 2D. Spectral analyses further confirmed that performance

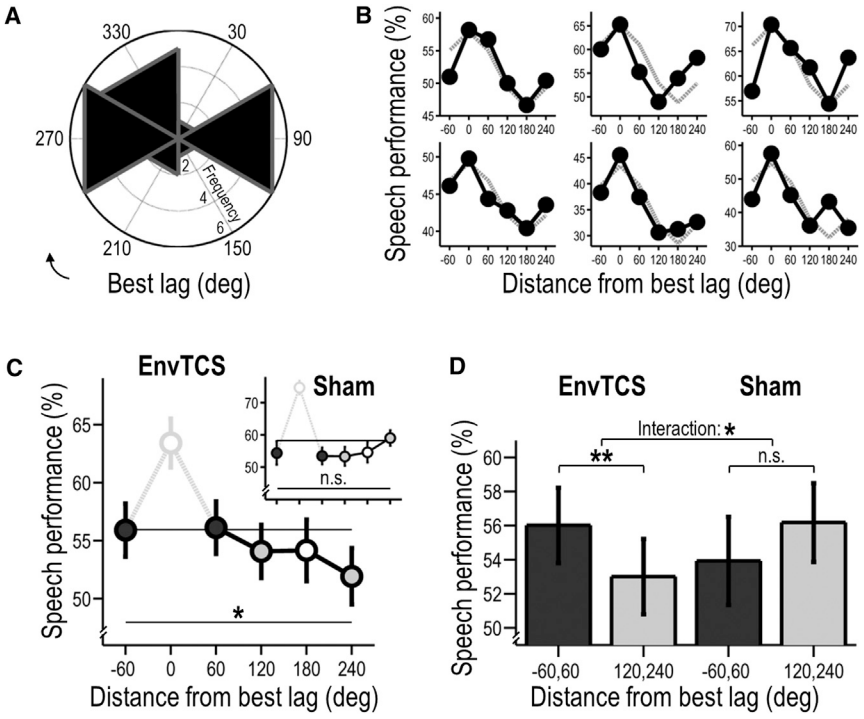

Figure 2. Results from the Two-Talker Experiment

(A) The phase angle histogram shows the distribution of listeners' best lag. This distribution did not deviate significantly from uniformity ($z = 0.68$, $p = 0.51$). On average, listeners' performance was best (63.4% ± 2.1%) when the aurally presented target envelope lagged behind envTCS by 316° ± 23°, which is equivalent to an audio lag of 219.5 ms or −30.5 ms given the cyclical nature of stimulation.

(B) Speech performance as a function of distance from best lag (i.e., presumed entrainment strength) for six exemplary listeners (black). Fitted f_c sinusoids (gray) are shown for reference to illustrate our initial predictions.

(C) Same as (B) but for the group (mean ± SEM across listeners), showing a main effect of best-lag distance (i.e., presumed entrainment strength) on speech performance. The peak performance at the best lag (0°) is trivial and was excluded from this analysis and (D). The horizontal line represents average overall performance under envTCS (55.9%). The inset shows analogously data from the control condition ("virtual-lag" sham stimulation; see STAR Methods). See also Figure S1A.

(D) Speech performance (mean ± SEM across listeners) averaged across the best-lag distances presumed to resemble an excitatory half-cycle (dark bar; see corresponding circle fillings in C) and inhibitory half-cycle (lighter bar) is shown for envTCS on the left and for the control condition on the right. Speech performance under envTCS was significantly better (on average 3.0 percentage points) during presumed excitatory versus inhibitory half-cycle, indicating that the temporal alignment between delta/theta neural excitability and auditory target-speech-evoked neural responses influenced intelligibility of the target talker. No such effect was observed in the control condition. See also Figure S2.

Corrected *$p < 0.05$, **$p < 0.005$; n.s., non-significant.

cycled most strongly in the delta/theta range (Figure S1A). Given that f_c TACS can entrain neural excitability and no other variable in our experiment cycled at f_c relative to the auditory target-speech rhythm, the most plausible explanation for the observed f_c cycle in speech performance is variations in speech-entrainment strength that were induced by the temporal (mis-)alignment between auditory target-speech-evoked neural responses and f_c fluctuations in neural excitability. These results provide evidence for a modulatory role of delta/theta speech entrainment in speech intelligibility in a two-talker setting.

Controls for Alternative Explanations

To exclude that the results above could reflect potential envTCS-unrelated influences, we conducted three control analyses. First, we applied the above analyses to data that were obtained under sham stimulation and stratified by (virtual) audio-lag condition (Figure 2C, inset). This did not replicate the observed effects on speech performance (no main effect of best-lag distance: $F_{4,80} = 1.28$, $p = 0.28$; no effect of presumed half-cycle: $t_{20} = -1.36$, $p = 0.91$); see Figures 2D and S1. A two-way ANOVA on speech performance, including stimulation condition (envTCS versus sham) and presumed half-cycle (excitatory versus inhibitory) as factors, revealed a significant dissociation (stimulation × half-cycle interaction: $F_{1,20} = 7.06$, $\eta^2 = 0.07$, corrected $p = 0.023$), further confirming that the observed effects did not occur spontaneously (i.e., without envTCS) but were caused specifically by envTCS. Second, we compared the virtual half-cycle difference (excitatory versus inhibitory half-cycle based on virtual-lag sham data) between participants who received sham stimulation during the first run of the experiment versus participants who received it in a later run. This revealed no significant difference (independent samples t test: $t_{19} = -0.051$, $p = 0.52$), suggesting that prior envTCS-induced phase entrainment did not carry over to later stimulation intervals. Finally, we analyzed participants' reported certainty of having received electric stimulation. This revealed no significant difference between envTCS runs versus sham runs ($t_{20} = 0.23$, $p = 0.82$), suggesting that participants were unaware of whether they received envTCS or sham stimulation.

Single-Talker Experiment

The subsequent "single-talker" experiment was designed to disambiguate whether the effect of speech-entrainment strength on intelligibility observed in the first experiment arose from changes in the perceptual separability of the auditory target stream or a more direct influence on intelligibility. It differed from the two-talker experiment in three key aspects: speech from individual talkers was presented in isolation, speech rhythm fluctuated naturally (i.e., stimulus rhythm was not periodic or fixed), and critical cues for speech entrainment or intelligibility in aural input were largely reduced. To implement this latter aspect, we artificially eliminated the critical speech rhythm from the aurally presented stimuli (thereby seriously hampering speech perception) and presented it via envTCS; see Figures 3A and 3B and Audio S2. In other words, the applied currents exactly matched

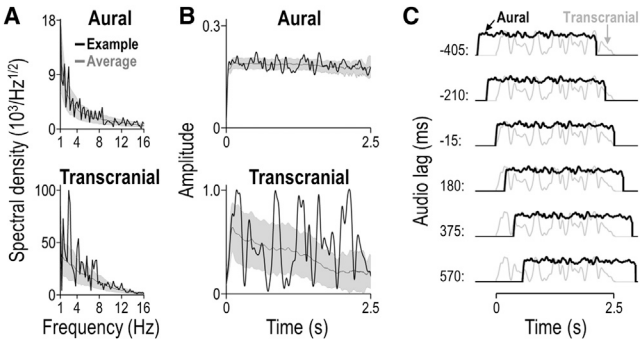

Figure 3. Stimulus Characteristics and Experimental Design for the Single-Talker Experiment

(A) Magnitude spectrum of the envelope of aurally presented degraded speech stimuli (top) and simultaneously applied transcranial current (envTCS, bottom). The black line represents a single exemplary sentence (same for top and bottom). The gray line and surrounding area represent summary statistics (mean ± SD) across all sentences. Note the clearer peaks, especially in the low-frequency range, for the exemplary envTCS spectrum (bottom, black line) and the lack of clear peaks in the average envTCS spectrum (bottom, gray line). These plots highlight that rhythmic cues for speech entrainment were carried primarily by envTCS, not the aurally presented stimuli, and that these rhythmic cues differed across sentences.

(B) Envelopes underlying the modulation spectra shown in (A). Note the near-flat envelope of the aurally presented stimuli (top) and the much larger fluctuations in envTCS (bottom). See also Audio S2.

(C) Sketch of the experimental design. As for the two-talker experiment, the six experimental conditions (rows) were characterized by the delay by which the aurally presented stimuli (black waveform; same as B, top) lagged behind envTCS (gray waveform; same as B, bottom).

the eliminated natural speech-envelope shapes and thereby could provide the listeners with perceptually relevant cues that were otherwise largely unavailable. This aspect was presumed to experimentally restore cortical entrainment to the degraded aural speech input. Analogous to the logic of the two-talker experiment, we parametrically varied the audio lag (Figure 3C) and predicted that this manipulation induces systematic changes in speech-entrainment strength that would be observable as corresponding changes in speech performance (for details, see STAR Methods). Alternatively, if speech entrainment has no direct influence on speech intelligibility, no change in behavioral performance should be observed.

To test this, we first obtained individual speech-benefit waveforms, which represent changes in speech performance induced by envTCS relative to a control condition (direct current stimulation devoid of any envelope information) as a function of audio lag. We compensated for individual differences in TCS-effect polarity by aligning the individual benefit waveforms based on their polarity (see STAR Methods). We tested whether the resulting waveforms exhibited systematic variations across audio lags.

Speech Entrainment Modulates Speech Intelligibility Even when Aural Speech-Amplitude Envelope Is Absent

Listeners correctly recognized on average 61.1% ± 1.7% of the words. The majority of listeners benefitted maximally from envTCS when aural input lagged behind envTCS by 375 ms (Figure 4A). Visual inspection of speech-benefit waveforms (Figures 4B and 4C) revealed that the maximum benefit was on average 4.7 ± 1.8 percentage points stronger than the minimum benefit; the latter was observed when the aural input led by 210 ms. Conforming to our predictions, a one-way ANOVA revealed that speech benefit varied significantly across audio lags ($F_{5,105}$ = 2.70, η^2 = 0.05, p = 0.025). Data from exploratory spectral analyses are shown in Figure S1B. Given that neural excitability follows quasi-rhythmic electric stimulation [24], the most plausible explanation for the observed change in speech benefit is variations in speech-entrainment strength that were induced by the temporal (mis-)alignment between auditory speech-evoked neural responses and envTCS-following neural excitability. This result provides evidence for an influence of speech entrainment on speech intelligibility in the absence of stream segregation.

Controls for Alternative Explanations

To exclude that the result above could reflect influences from potential envTCS-induced tactile cues, we correlated participants' speech-benefit range (maximum versus minimum) with their reported amount of attention paid to the electric stimulation. This revealed no significant association (Kendall's $\tau = -0.29$, p = 0.10), suggesting that participants could not exploit potential tactile cues for performing the task.

DISCUSSION

Our results show consistently an effect of envTCS timing on speech-recognition performance in a cocktail-party-like scenario and a listening situation devoid of aural speech-amplitude envelope input (two-talker and single-talker experiment, respectively). Given that transcranial current stimulation entrains cortical excitability (see Introduction), these findings reveal a causal role of low-frequency neural speech entrainment in speech intelligibility.

Speech Entrainment Modulates Intelligibility

Our finding of a causal role of speech entrainment in speech intelligibility is novel. Current neurolinguistic models explain speech perception based on slow neural excitability fluctuations [1, 16–21], and computational models of auditory cortical speech analysis have provided support for this notion [31]; however, empirical evidence was still missing. Previous speech studies using non-invasive brain stimulation found that TACS at 40 Hz, but not 6 Hz, impairs perceptual learning of a phoneme-categorization task [32]. However, speech entrainment and its effect on intelligibility could not be assessed. Studies using neuroelectromagnetic recordings observed associations between speech entrainment and intelligibility (see Introduction), and these associations seem to depend on attention [9]. However, manipulations of intelligibility were achieved by varying the acoustic stimuli (e.g., via time compression/reversal, e.g., [5, 23], or spectral degradation/masking, e.g., [12, 22]), leaving unclear whether the observed effects on entrainment were caused by changes in acoustic input, intelligibility, or both. By exploiting the modulatory ability of envTCS, our study could isolate an effect of speech entrainment on intelligibility without concomitant acoustic changes and therewith disambiguate the previous results.

Speech Entrainment and Intelligibility Interact Reciprocally

Only few studies could establish an association between entrainment and intelligibility in the absence of acoustic confounding.

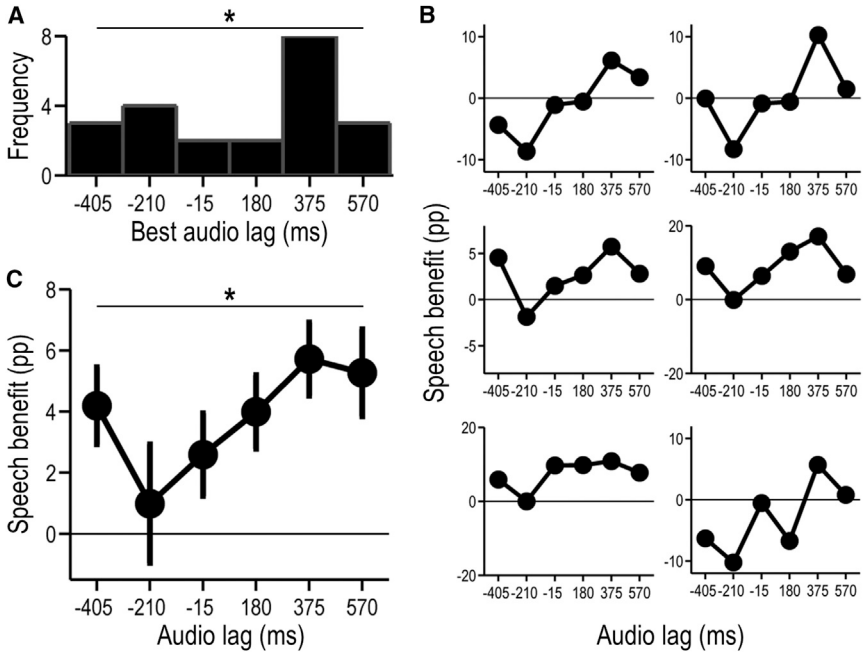

Figure 4. Results from the Single-Talker Experiment

(A) Histogram of listeners' best lag after alignment. The distribution reveals significant concentration ($D_{22,100}$ = 0.31, p = 0.049) on the condition where envTCS preceded aural input by 375 ms. See also Figure S4.

(B) Individual speech-benefit waveforms for six exemplary listeners. The waveforms illustrate the benefit from envTCS for listeners' speech performance (expressed in units of percentage points, pp) as a function of audio lag.

(C) Same as (B) but for the group (mean ± SEM across listeners), revealing a main effect of audio lag on speech benefit. The overall magnitude of benefit could be biased due to the alignment. *p < 0.05. See also Figure S1B.

However, the quasi-experimental nature of the applied study designs hampered causal inference [8, 33–35] and overall results have been mixed [36, 37]. Evidence for a functional contribution of intelligibility to cortical entrainment comes from studies showing that shuffling the order of syllables in a sequence of trisyllabic words reduces cortical entrainment to these words [38]. Similarly, it has been shown that shuffling words in a sequence of sentences preserves cortical entrainment to the words, but not to the sentences [8]. In line with this, our results show a positive relation between entrainment and intelligibility but with reversed causal direction. Together, these findings highlight that entrainment and intelligibility interact reciprocally: speech entrainment modulates the intelligibility of individual words in sentences (present study) and word comprehension may further enhance this entrainment [8, 38]. From this "bidirectional" perspective, the previously observed acoustic effects on intelligibility were mediated in bottom-up fashion by speech entrainment, whereas those on entrainment were partially mediated top-down by speech comprehension. Although our results indicate that entrainment contributes critically to intelligibility, they do not imply that entrainment is sufficient for intelligibility nor that it requires intelligibility, given that entrainment also occurs for unintelligible (e.g., time-reversed) speech sounds and does not render these sounds more intelligible (e.g., [22, 37]).

Which Processes Underlie the Observed envTCS Effects on Speech-Recognition Performance?

Our findings provide fundamental insights into mechanisms underlying auditory speech comprehension. Neural entrainment aligns excitability to a temporal pattern (e.g., [26]), which provides a mechanism for identifying that specific pattern (and possibly nested patterns) in upcoming sensory input (e.g., [15]). The relevant temporal pattern is derived from salient temporal structures in external stimuli (e.g., [5]) and/or informative temporal structures that the perceiver expects (or focuses at) based on prior knowledge [8, 9, 34, 39]. Referring to these driving forces as "exogenous entrainers" and "endogenous entrainers," respectively, we interpret our observation of audio-lag-induced changes in speech-recognition performance as follows: our experiments involved an electric exogenous entrainer (envTCS) and presumably a linguistic endogenous entrainer (listeners' expectancy of linguistic structures). Depending on the audio lag, the electric entrainer attracted any linguistic entrainer (in terms of shape and timing of the expected temporal structure) toward or away from auditory-evoked temporal response patterns representing intelligibility-relevant information. This facilitated or hampered identification and processing of this critical information, leading to the observed speech-performance changes. Put differently, we interpret our results as arising from an entrainment-based process for identifying perceptually relevant temporal response patterns.

We cannot disentangle whether neural excitability in our study followed primarily the amplitude of envTCS or temporally correlated features that further support cortical speech entrainment and/or intelligibility (if presented aurally), such as acoustic edges or phoneme borders. These possibilities may be disambiguated in the future by comparing envTCS-induced intelligibility modulations for different current shapes that emphasize these different speech-envelope features. Similarly, the purely behavioral nature of our measures, the poor spatial specificity of TCS, and the currently still limited mechanistic understanding of this technique in the living human brain [40, 41] do not allow us to disentangle whether the observed effects arise from low-level general-purpose auditory processes (e.g., sensory gain modulation), higher-level speech-specific processes (e.g., lexical-semantic pattern analysis), or both (cf. [16, 21]). A contribution from generic (speech-unspecific) central auditory processes is supported by (1) the fact that our two-talker results closely resemble results from a matching TACS study with non-speech sounds [42], (2) findings from other non-speech-entrainment studies [43], and (3) no evidence for TACS effects on peripheral auditory processing [44]. This notion may be tested in the future by combining envTCS with simultaneous neuroimaging or electrophysiology [45] and directly comparing envTCS-induced

neural response modulations for speech stimuli versus non-speech equivalents [46].

Given the cyclical nature of envTCS in the two-talker experiment, the observed cycle in speech performance reflects modulations induced by envTCS-entrained delta/theta neural oscillations. For the single-talker experiment, such an interpretation in terms of envTCS-entrained oscillations is hampered by two factors: although envTCS fluctuated markedly in the delta/theta range (black line in Figure 3A, bottom), its spectral profile varied non-systematically across sentences (illustrated by the lack of a clear peak in the average envelope spectrum in Figure 3A, bottom). Moreover, the (time domain) audio-lag manipulation induced variable phase shifts across oscillatory frequencies and thus could not induce coherent variations in oscillatory speech-entrainment strength. This implies that our observed effect on speech benefit most likely reflects modulation of auditory speech-evoked responses by *non-oscillatory* envTCS-following neural excitability changes. Nevertheless, we note that the effect exhibits the shape of a long cycle (~0.9 Hz; Figures S1B and 4C), which suggests that onsets of envTCS (at the beginning of trials) reset the phase of low-delta neural oscillations. Given that slow endogenous cortical activity entrains to perceived linguistic structures of correspondingly long timescales [8], it is conceivable that the presumed envTCS onset-induced phase resets initiated analysis for such large linguistic structures, which remains an idea to be tested in future studies.

A conceptual advantage of the non-cyclical and trial-specific nature of envTCS in our single-talker experiment is that it circumvents confusion of consecutive cycles, thus providing a clearer picture of the timing of the underlying processes. Our observation that most listeners took the strongest advantage of the transcranial speech cues when these *preceded* the aural input hints at neural processes involved in temporal prediction. The observation of the best audio lag at 375 ms is reminiscent of findings from audiovisual speech studies that have shown that intelligibility remains stable for audio lags of up to ~240 ms [47]. Taking into account that acoustic speech envelope follows articulatory facial movements with a delay of up to ~200 ms (e.g., [48]) and that neural effects of electric stimulation have probably shorter latencies (~40 ms [25]) than visual input to auditory cortex (~100 ms [49]), our single-talker envTCS results may reflect operation of a similar mechanism in auditory cortex as in audiovisual speech recognition. Indeed, articulatory facial cues support auditory cortical speech entrainment and thereby facilitate processing of later-arriving acoustic speech input (e.g., [49, 50]), which indicates that these visual temporal cues play a predictive role in audiovisual speech processing (e.g., [51]). Our observation of the worst lag at −210 ms suggests that transcranial speech cues may influence neural speech parsing not only by facilitating the processing of upcoming auditory speech input (+375 ms; see above) but also by hampering processing of recent input (−210 ms).

In sum, our results support the view that neural excitability follows the temporal pattern of envTCS and thereby modulates the processing of corresponding temporal structures in subsequent acoustic input, which appears to be critical for speech intelligibility (in the case of linguistic temporal structures). Whether onsets of envTCS are sufficient to interfere with listeners' expectancies of large linguistic structures remains to be investigated in future studies.

Applicability of envTCS

Beyond these theoretical implications, our findings may pave the way for interesting practical implications. A novel aspect of our single-talker experiment is that the applied brain stimulation featured more complex temporal patterns than conventionally applied constant, alternating, or random stimulation [52]. Our finding that this complex-shaped stimulation influences listeners' ability to decipher degraded auditory speech suggests that our approach can be utilized to inform neural speech analysis about specific quasi-rhythmic linguistic features in upcoming speech input, such as words, phrases, or sentences [8]. Because we observed effects of primarily suppressive and predictive nature (hampering performance by approximately 5 percentage points; Figure S2), envTCS could primarily serve the suppression of predefined speech-input features. More generally, it could be utilized as a "transcranial transmitter" of complex, behaviorally relevant temporal information to subliminally control a perceiver's temporal attention in any sensory modality (vision, audition, or touch). It might further serve rhythmic training-based interventions of developmental dyslexia [53], a highly prevalent reading/spelling disorder associated with alterations in acoustic input-driven cortical entrainment. To enable such envisioned applications, future research needs to systematically optimize envTCS parameters (e.g., the number and positions of electric stimulation channels) in normal and clinical populations.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- CONTACT FOR REAGENT AND RESOURCE SHARING
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Participants
- METHOD DETAILS
  - Auditory stimulation
  - Electric stimulation
  - Auditory-electric stimulus presentation
  - Task and experimental design
  - Procedure
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Data analysis
- DATA AND SOFTWARE AVAILABILITY

SUPPLEMENTAL INFORMATION

Supplemental Information includes four figures and two audio files and can be found with this article online at https://doi.org/10.1016/j.cub.2017.11.033.

ACKNOWLEDGMENTS

We thank Jules Erkens, Mahan Hosseini, Andrei Razoare, Rutger Slegers, and Nassim Sedaghat for help with conducting the study. We thank Anne Kösem and four anonymous reviewers for useful comments on the manuscript. This work was supported by the Netherlands Organization for Scientific Research (Vici 918-17-603 to D.B. and Vici 435-12-002 to E.F.) and the French National Research Agency (LabEx CeLyA; ANR-10-LABX-0060/ANR-11-IDEX-0007).

AUTHOR CONTRIBUTIONS

Conceptualization, L.R.; Methodology, L.R., D.B., E.G., and B.S.; Software, L.R., E.G., and D.B.; Investigation, L.R.; Resources, L.R. and E.F.; Writing – Original Draft, L.R.; Writing – Review & Editing, L.R., D.B., E.G., B.S., and E.F.

Received: September 14, 2017
Revised: October 26, 2017
Accepted: November 15, 2017
Published: December 28, 2017

REFERENCES

1. Zoefel, B., and VanRullen, R. (2015). The role of high-level processes for oscillatory phase entrainment to speech sound. Front. Hum. Neurosci. *9*, 651.
2. Ding, N., Patel, A.D., Chen, L., Butler, H., Luo, C., and Poeppel, D. (2017). Temporal modulations in speech and music. Neurosci. Biobehav. Rev. Published online February 14, 2017. https://doi.org/10.1016/j.neubiorev.2017.02.011.
3. Ghitza, O. (2012). On the role of theta-driven syllabic parsing in decoding speech: intelligibility of speech with a manipulated modulation spectrum. Front. Psychol. *3*, 238.
4. Drullman, R., Festen, J.M., and Plomp, R. (1994). Effect of reducing slow temporal modulations on speech reception. J. Acoust. Soc. Am. *95*, 2670–2680.
5. Gross, J., Hoogenboom, N., Thut, G., Schyns, P., Panzeri, S., Belin, P., and Garrod, S. (2013). Speech rhythms and multiplexed oscillatory sensory coding in the human brain. PLoS Biol. *11*, e1001752.
6. Zoefel, B., and VanRullen, R. (2015). Selective perceptual phase entrainment to speech rhythm in the absence of spectral energy fluctuations. J. Neurosci. *35*, 1954–1964.
7. Di Liberto, G.M., O'Sullivan, J.A., and Lalor, E.C. (2015). Low-frequency cortical entrainment to speech reflects phoneme-level processing. Curr. Biol. *25*, 2457–2465.
8. Ding, N., Melloni, L., Zhang, H., Tian, X., and Poeppel, D. (2016). Cortical tracking of hierarchical linguistic structures in connected speech. Nat. Neurosci. *19*, 158–164.
9. Rimmele, J.M., Zion Golumbic, E., Schröger, E., and Poeppel, D. (2015). The effects of selective attention and speech acoustics on neural speech-tracking in a multi-talker scene. Cortex *68*, 144–154.
10. Zion Golumbic, E.M., Ding, N., Bickel, S., Lakatos, P., Schevon, C.A., McKhann, G.M., Goodman, R.R., Emerson, R., Mehta, A.D., Simon, J.Z., et al. (2013). Mechanisms underlying selective neuronal tracking of attended speech at a "cocktail party". Neuron *77*, 980–991.
11. Horton, C., D'Zmura, M., and Srinivasan, R. (2013). Suppression of competing speech through entrainment of cortical oscillations. J. Neurophysiol. *109*, 3082–3093.
12. Ding, N., and Simon, J.Z. (2012). Emergence of neural encoding of auditory objects while listening to competing speakers. Proc. Natl. Acad. Sci. USA *109*, 11854–11859.
13. Kayser, S.J., Ince, R.A., Gross, J., and Kayser, C. (2015). Irregular speech rate dissociates auditory cortical entrainment, evoked responses, and frontal alpha. J. Neurosci. *35*, 14691–14701.
14. Mesgarani, N., and Chang, E.F. (2012). Selective cortical representation of attended speaker in multi-talker speech perception. Nature *485*, 233–236.
15. Schroeder, C.E., and Lakatos, P. (2009). Low-frequency neuronal oscillations as instruments of sensory selection. Trends Neurosci. *32*, 9–18.
16. Giraud, A.L., and Poeppel, D. (2012). Cortical oscillations and speech processing: emerging computational principles and operations. Nat. Neurosci. *15*, 511–517.
17. Peelle, J.E., and Davis, M.H. (2012). Neural oscillations carry speech rhythm through to comprehension. Front. Psychol. *3*, 320.
18. Ghitza, O. (2011). Linking speech perception and neurophysiology: speech decoding guided by cascaded oscillators locked to the input rhythm. Front. Psychol. *2*, 130.
19. Ding, N., and Simon, J.Z. (2014). Cortical entrainment to continuous speech: functional roles and interpretations. Front. Hum. Neurosci. *8*, 311.
20. Zion Golumbic, E.M., Poeppel, D., and Schroeder, C.E. (2012). Temporal context in speech processing and attentional stream selection: a behavioral and neural perspective. Brain Lang. *122*, 151–161.
21. Kösem, A., and van Wassenhove, V. (2016). Distinct contributions of low- and high-frequency neural oscillations to speech comprehension. Lang. Cogn. Neurosci. *32*, 536–544.
22. Peelle, J.E., Gross, J., and Davis, M.H. (2013). Phase-locked responses to speech in human auditory cortex are enhanced during comprehension. Cereb. Cortex *23*, 1378–1387.
23. Ahissar, E., Nagarajan, S., Ahissar, M., Protopapas, A., Mahncke, H., and Merzenich, M.M. (2001). Speech comprehension is correlated with temporal response patterns recorded from auditory cortex. Proc. Natl. Acad. Sci. USA *98*, 13367–13372.
24. Fröhlich, F., and McCormick, D.A. (2010). Endogenous electric fields may guide neocortical network activity. Neuron *67*, 129–143.
25. Bikson, M., Inoue, M., Akiyama, H., Deans, J.K., Fox, J.E., Miyakawa, H., and Jefferys, J.G. (2004). Effects of uniform extracellular DC electric fields on excitability in rat hippocampal slices in vitro. J. Physiol. *557*, 175–190.
26. Kayser, C., Wilson, C., Safaai, H., Sakata, S., and Panzeri, S. (2015). Rhythmic auditory cortex activity at multiple timescales shapes stimulus-response gain and background firing. J. Neurosci. *35*, 7750–7762.
27. Riecke, L., Formisano, E., Herrmann, C.S., and Sack, A.T. (2015). 4-Hz transcranial alternating current stimulation phase modulates hearing. Brain Stimul. *8*, 777–783.
28. Nourski, K.V., Reale, R.A., Oya, H., Kawasaki, H., Kovach, C.K., Chen, H., Howard, M.A., 3rd, and Brugge, J.F. (2009). Temporal envelope of time-compressed speech represented in the human auditory cortex. J. Neurosci. *29*, 15564–15574.
29. Kubanek, J., Brunner, P., Gunduz, A., Poeppel, D., and Schalk, G. (2013). The tracking of speech envelope in the human cortex. PLoS ONE *8*, e53398.
30. Ali, M.M., Sellers, K.K., and Fröhlich, F. (2013). Transcranial alternating current stimulation modulates large-scale cortical network activity by network resonance. J. Neurosci. *33*, 11262–11275.
31. Hyafil, A., Fontolan, L., Kabdebon, C., Gutkin, B., and Giraud, A.L. (2015). Speech encoding by coupled cortical theta and gamma oscillations. eLife *4*, e06213.
32. Rufener, K.S., Zaehle, T., Oechslin, M.S., and Meyer, M. (2016). 40Hz-transcranial alternating current stimulation (tACS) selectively modulates speech perception. Int. J. Psychophysiol. *101*, 18–24.
33. Meyer, L., Henry, M.J., Gaston, P., Schmuck, N., and Friederici, A.D. (2017). Linguistic bias modulates interpretation of speech via neural delta-band oscillations. Cereb. Cortex *27*, 4293–4302.
34. Kösem, A., Basirat, A., Azizi, L., and van Wassenhove, V. (2016). High-frequency neural activity predicts word parsing in ambiguous speech streams. J. Neurophysiol. *116*, 2497–2512.
35. Pérez, A., Carreiras, M., Gillon Dowens, M., and Duñabeitia, J.A. (2015). Differential oscillatory encoding of foreign speech. Brain Lang. *147*, 51–57.
36. Peña, M., and Melloni, L. (2012). Brain oscillations during spoken sentence processing. J. Cogn. Neurosci. *24*, 1149–1164.
37. Millman, R.E., Johnson, S.R., and Prendergast, G. (2015). The role of phase-locking to the temporal envelope of speech in auditory perception and speech intelligibility. J. Cogn. Neurosci. *27*, 533–545.
38. Buiatti, M., Peña, M., and Dehaene-Lambertz, G. (2009). Investigating the neural correlates of continuous speech computation with frequency-tagged neuroelectric responses. Neuroimage *44*, 509–519.

39. Stefanics, G., Hangya, B., Hernádi, I., Winkler, I., Lakatos, P., and Ulbert, I. (2010). Phase entrainment of human delta oscillations can mediate the effects of expectation on reaction speed. J. Neurosci. *30*, 13578–13585.

40. Lafon, B., Henin, S., Huang, Y., Friedman, D., Melloni, L., Thesen, T., Doyle, W., Buzsáki, G., Devinsky, O., Parra, L.C., and A Liu, A. (2017). Low frequency transcranial electrical stimulation does not entrain sleep rhythms measured by human intracranial recordings. Nat. Commun. *8*, 1199.

41. Opitz, A., Falchier, A., Yan, C.G., Yeagle, E.M., Linn, G.S., Megevand, P., Thielscher, A., Deborah A, R., Milham, M.P., Mehta, A.D., and Schroeder, C.E. (2016). Spatiotemporal structure of intracranial electric fields induced by transcranial electric stimulation in humans and nonhuman primates. Sci. Rep. *6*, 31236.

42. Riecke, L., Sack, A.T., and Schroeder, C.E. (2015). Endogenous delta/theta sound-brain phase entrainment accelerates the buildup of auditory streaming. Curr. Biol. *25*, 3196–3201.

43. Doelling, K.B., and Poeppel, D. (2015). Cortical entrainment to music and its modulation by expertise. Proc. Natl. Acad. Sci. USA *112*, E6233–E6242.

44. Ueberfuhr, M.A., Braun, A., Wiegrebe, L., Grothe, B., and Drexl, M. (2017). Modulation of auditory percepts by transcutaneous electrical stimulation. Hear. Res. *350*, 235–243.

45. Bergmann, T.O., Karabanov, A., Hartwigsen, G., Thielscher, A., and Siebner, H.R. (2016). Combining non-invasive transcranial brain stimulation with neuroimaging and electrophysiology: current approaches and future perspectives. Neuroimage *140*, 4–19.

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

47. Grant, K.W., van Wassenhove, V., and Poeppel, D. (2004). Detection of auditory (cross-spectral) and auditory-visual (cross-modal) synchrony. Speech Commun. *44*, 43–53.

48. Schwartz, J.L., and Savariaux, C. (2014). No, there is no 150 ms lead of visual speech on auditory speech, but a range of audiovisual asynchronies varying from small audio lead to large audio lag. PLoS Comput. Biol. *10*, e1003743.

49. Schroeder, C.E., Lakatos, P., Kajikawa, Y., Partan, S., and Puce, A. (2008). Neuronal oscillations and visual amplification of speech. Trends Cogn. Sci. *12*, 106–113.

50. Crosse, M.J., Butler, J.S., and Lalor, E.C. (2015). Congruent visual speech enhances cortical entrainment to continuous auditory speech in noise-free conditions. J. Neurosci. *35*, 14195–14204.

51. Peelle, J.E., and Sommers, M.S. (2015). Prediction and constraint in audiovisual speech perception. Cortex *68*, 169–181.

52. Paulus, W. (2011). Transcranial electrical stimulation (tES - tDCS; tRNS, tACS) methods. Neuropsychol. Rehabil. *21*, 602–617.

53. Bhide, A., Power, A., and Goswami, U. (2013). A rhythmic musical intervention for poor readers: a comparison of efficacy with a letter-based intervention. Mind Brain Educ. *7*, 113–123.

54. Versfeld, N.J., Daalder, L., Festen, J.M., and Houtgast, T. (2000). Method for the selection of sentence materials for efficient measurement of the speech reception threshold. J. Acoust. Soc. Am. *107*, 1671–1684.

55. Boersma, P. (2001). Praat, a system for doing phonetics by computer. Glot Int. *5*, 341–345.

56. Ten Oever, S., de Graaf, T.A., Bonnemayer, C., Ronner, J., Sack, A.T., and Riecke, L. (2016). Stimulus presentation at specific neuronal oscillatory phases experimentally controlled with tACS: implementation and applications. Front. Cell. Neurosci. *10*, 240.

57. Van Noorden, L.P.A.S. (1975). Temporal Coherence in the Perception of Tone Sequences (University of Technology).

58. Gaudrain, E., Grimault, N., Healy, E.W., and Béra, J.C. (2008). Streaming of vowel sequences based on fundamental frequency in a cochlear-implant simulation. J. Acoust. Soc. Am. *124*, 3076–3087.

59. Moulines, E., and Charpentier, F. (1990). Pitch-synchronous waveform processing techniques for text-to-speech synthesis using diphones. Speech Commun. *9*, 453–467.

60. Verhoeven, J., De Pauw, G., and Kloots, H. (2004). Speech rate in a pluricentric language: a comparison between Dutch in Belgium and the Netherlands. Lang. Speech *47*, 297–308.

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

62. Moore, B.C.J. (2003). An Introduction to the Psychology of Hearing, Fifth Edition (Academic Press).

63. Drullman, R., Festen, J.M., and Plomp, R. (1994). Effect of temporal envelope smearing on speech reception. J. Acoust. Soc. Am. *95*, 1053–1064.

64. Füllgrabe, C., Stone, M.A., and Moore, B.C. (2009). Contribution of very low amplitude-modulation rates to intelligibility in a competing-speech task (L). J. Acoust. Soc. Am. *125*, 1277–1280.

65. Chait, M., Greenberg, S., Arai, T., Simon, J.Z., and Poeppel, D. (2015). Multi-time resolution analysis of speech: evidence from psychophysics. Front. Neurosci. *9*, 214.

66. Ghitza, O. (2001). On the upper cutoff frequency of the auditory critical-band envelope detectors in the context of speech perception. J. Acoust. Soc. Am. *110*, 1628–1640.

67. Gilbert, G., and Lorenzi, C. (2006). The ability of listeners to use recovered envelope cues from speech fine structure. J. Acoust. Soc. Am. *119*, 2438–2444.

68. Zeng, F.G., Nie, K., Liu, S., Stickney, G., Del Rio, E., Kong, Y.Y., and Chen, H. (2004). On the dichotomy in auditory perception between temporal envelope and fine structure cues. J. Acoust. Soc. Am. *116*, 1351–1354.

69. Zoefel, B., and VanRullen, R. (2016). EEG oscillations entrain their phase to high-level features of speech sound. Neuroimage *124* (Pt A), 16–23.

70. Moon, I.J., Won, J.H., Park, M.H., Ives, D.T., Nie, K., Heinz, M.G., Lorenzi, C., and Rubinstein, J.T. (2014). Optimal combination of neural temporal envelope and fine structure cues to explain speech identification in background noise. J. Neurosci. *34*, 12145–12154.

71. Heimrath, K., Fiene, M., Rufener, K.S., and Zaehle, T. (2016). Modulating human auditory processing by transcranial electrical stimulation. Front. Cell. Neurosci. *10*, 53.

72. Zion Golumbic, E., Cogan, G.B., Schroeder, C.E., and Poeppel, D. (2013). Visual input enhances selective speech envelope tracking in auditory cortex at a "cocktail party". J. Neurosci. *33*, 1417–1426.

73. Luo, H., Liu, Z., and Poeppel, D. (2010). Auditory cortex tracks both auditory and visual stimulus dynamics using low-frequency neuronal phase modulation. PLoS Biol. *8*, e1000445.

74. Park, H., Kayser, C., Thut, G., and Gross, J. (2016). Lip movements entrain the observers' low-frequency brain oscillations to facilitate speech intelligibility. eLife *5*, e14521.

75. Stonkus, R., Braun, V., Kerlin, J.R., Volberg, G., and Hanslmayr, S. (2016). Probing the causal role of prestimulus interregional synchrony for perceptual integration via tACS. Sci. Rep. *6*, 32065.

76. Kar, K., and Krekelberg, B. (2014). Transcranial alternating current stimulation attenuates visual motion adaptation. J. Neurosci. *34*, 7334–7340.

77. Sohoglu, E., Peelle, J.E., Carlyon, R.P., and Davis, M.H. (2012). Predictive top-down integration of prior knowledge during speech perception. J. Neurosci. *32*, 8443–8453.

78. Tuennerhoff, J., and Noppeney, U. (2016). When sentences live up to your expectations. Neuroimage *124* (Pt A), 641–653.

79. Neuling, T., Rach, S., Wagner, S., Wolters, C.H., and Herrmann, C.S. (2012). Good vibrations: oscillatory phase shapes perception. Neuroimage *63*, 771–778.

80. Marshall, L., Helgadóttir, H., Mölle, M., and Born, J. (2006). Boosting slow oscillations during sleep potentiates memory. Nature *444*, 610–613.

81. Lakatos, P., O'Connell, M.N., Barczak, A., Mills, A., Javitt, D.C., and Schroeder, C.E. (2009). The leading sense: supramodal control of neurophysiological context by attention. Neuron *64*, 419–430.

82. van Wassenhove, V., Grant, K.W., and Poeppel, D. (2007). Temporal window of integration in auditory-visual speech perception. Neuropsychologia *45*, 598–607.

83. Thorne, J.D., and Debener, S. (2014). Look now and hear what's coming: on the functional role of cross-modal phase reset. Hear. Res. *307*, 144–152.

84. Maier, J.X., Di Luca, M., and Noppeney, U. (2011). Audiovisual asynchrony detection in human speech. J. Exp. Psychol. Hum. Percept. Perform. *37*, 245–256.

85. Schroeder, C.E., and Foxe, J.J. (2002). The timing and laminar profile of converging inputs to multisensory areas of the macaque neocortex. Brain Res. Cogn. Brain Res. *14*, 187–198.

86. Krause, B., and Cohen Kadosh, R. (2014). Not all brains are created equal: the relevance of individual differences in responsiveness to transcranial electrical stimulation. Front. Syst. Neurosci. *8*, 25.

87. Kabakov, A.Y., Muller, P.A., Pascual-Leone, A., Jensen, F.E., and Rotenberg, A. (2012). Contribution of axonal orientation to pathway-dependent modulation of excitatory transmission by direct current stimulation in isolated rat hippocampus. J. Neurophysiol. *107*, 1881–1889.

88. Campain, R., and Minckler, J. (1976). A note on the gross configurations of the human auditory cortex. Brain Lang. *3*, 318–323.

89. Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate - a practical and powerful approach to multiple testing. J. R. Stat. Soc. B *57*, 289–300.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited Data		
Stimulus generation code	This paper; Mendeley Data	https://doi.org/10.17632/zwd67kjpdd.1
Single-subject results	This paper; Mendeley Data	https://doi.org/10.17632/zwd67kjpdd.1
Software and Algorithms		
Presentation	Neurobehavioral Systems; this paper	https://www.neurobs.com/
MATLAB	MathWorks	https://www.mathworks.com/
Speech materials	[54]	https://www.vumc.nl/afdelingen/kno/
Praat	[55]	http://www.fon.hum.uva.nl/praat/
Datastreamer	[56]	https://osf.io/h6b8v/

CONTACT FOR REAGENT AND RESOURCE SHARING

Information and requests for resources should be directed to and will be fulfilled by the Lead Contact, Lars Riecke (l.riecke@maastrichtuniversity.nl). Usage of the speech materials ('zinsmateriaal VU98') is subject to license agreement with VU University Medical Center, Amsterdam, the Netherlands [54].

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Participants

Twenty-two native Dutch volunteers (11 females, ages: 18–28 years) participated in both experiments. They reported no history of neurological, psychiatric, or hearing disorders, were suited to undergo TCS as assessed by prior screening, and gave their written informed consent before taking part. They had normal hearing (defined as hearing thresholds $\leq$ 25dB HL at 0.25, 0.5, 1, 2, 4, and 6kHz), except for one participant who had mild hearing loss in the right ear for 0.5kHz (excluding this participant's data from the analyses did not alter the results qualitatively, i.e., not changing the conclusions that can be drawn from the study). One participant's data had to be excluded from the two-talker experiment due to a technical problem during the data acquisition. Participants received study credits or monetary reward for their participation. The experimental procedure was approved by the local research ethics committee (*Ethical Review Committee Psychology and Neuroscience,* Maastricht University).

METHOD DETAILS

Auditory stimulation

Speech stimuli were generated from a corpus of 1014 meaningful everyday Dutch sentences [54]. Each sentence consisted of a total number of eight or nine syllables distributed over four to nine words and each word consisted of maximally three syllables. Half of the sentences were spoken by a male talker and the other half by a female talker.

Two-talker experiment

Inspired by a classical auditory streaming paradigm with alternating sound sequences [57, 58], we designed two-talker auditory stimuli with the aim to alternate the talkers' syllables at a strong, fixed rhythm. Synthesis of the stimuli involved the following steps: First, the rms level was fixed across recordings. Second, average syllable rate was fixed to f_c across recordings by temporally compressing or expanding the recordings without altering the voice pitch. This was done using the pitch-synchronous overlap-add method [59] as implemented in PRAAT software [55]. As a result, phases containing little phonetic content (e.g., silences between syllables) alternated with phases containing strong phonetic content at an average rate f_c. Third, this rhythm was enhanced by applying an f_c-sinusoidal amplitude modulation (depth: 50%, same phase as the f_c envelope of the original signal). Informal listening tests confirmed that these processing steps did not noticeably hamper the intelligibility and naturalness of the individual sentences. Fourth, the preprocessed recordings were ranked according to how well their envelope resembled the envTCS current (described below) as quantified by cross-correlation. Fifth, the 338 best-fitting sentences from each talker were selected, ensuring a matched proportion of eight- and nine-syllable sentences across talkers. Sixth, the selected recordings were temporally aligned with respect to their f_c envelopes, the female talker was delayed by an f_c half-cycle, and any silent f_c cycle at the beginning of any recording was discarded. Finally, recordings from the two talkers were mixed while minimizing the overall between-talker difference in audio-onset time; this

was done separately for eight- and nine-syllable sentences. Thus, the individual sentences in the resulting two-talker stimuli largely overlapped in time while their constituent syllables alternated at $2f_c$. An exemplary stimulus is provided in Audio S1.

We chose the critical stimulation frequency f_c to be 4Hz because it closely matches the speech rate of the original speech corpus (on average 4.4Hz) and the average syllable rate in Dutch [60], it has been associated with cortical syllable tracking [16, 61], and cortical oscillations at this frequency can be entrained with transcranial alternating current stimulation [27, 42].

To control speech intelligibility, we added noise to the two-talker stimuli and varied the signal-to-noise ratio (SNR) by adjusting the noise level. The noise was stationary and its spectrum was shaped to match the average power spectrum of the two-talker stimuli. The noise extended before and after the two-talker stimulus by one f_c period, respectively, including long 250-ms raised-cosine ramps to reduce potential auditory-evoked neural phase resetting. The final combined auditory stimuli were presented at an average sound level of 60dB SPL.

Single-talker experiment

The single-talker auditory stimuli were designed to carry no aural amplitude-envelope cues for speech entrainment and intelligibility. These stimuli were synthesized using vocoders as follows: First, the recordings were passed through a bank of 30 gamma-tone filters with center frequencies ranging from 87Hz to 6930Hz equi-spaced on a Cam scale (see [62] for details of this unit). Second, the temporal-fine structure and the amplitude envelope were extracted from each channel signal using the Hilbert transform. Third, each channel envelope was further decomposed into a low- and high-frequency portion (4^{th}-order Butterworth filters, cutoff frequencies: 16Hz lowpass and 64Hz highpass), respectively. The low-frequency portion, which carries salient cues for speech entrainment and intelligibility, was used later to define the envTCS current (described below). Fourth, the high-frequency portion was multiplied with the channel temporal-fine structure. Fifth, the resulting channel signal was summed across channels to synthesize the vocoded speech signal, which thus excluded the low-frequency envelopes. Finally, rms level was fixed and 25-ms onset and offset ramps were applied. The stimuli were presented at an average sound level of 67dB SPL. An exemplary stimulus is provided in Audio S2.

We chose the aforementioned filter-cutoff frequencies because narrow-band speech-envelope fluctuations below 16Hz, but not above 64Hz, contribute strongly to speech intelligibility [4, 63–65]. Moreover, speech-envelope fluctuations below 16Hz strongly contribute to cortical speech entrainment (see Introduction). We chose the specific filter-bank settings because they reduce possible recovery of speech-envelope cues in the peripheral auditory system [66–68]. Although all these stimulus modifications substantially reduce aural entrainment cues (i.e., amplitude envelope), they cannot completely abolish temporally correlated higher-order acoustic or linguistic features (e.g., phonetic information) [6]. Therefore, our envelope-reduced stimuli may elicit overall reduced envelope-following responses, with primary contributions from neuronal populations tuned to the residual features [69].

To control listeners' speech-performance level, we scaled the high-frequency envelope of the channel signals (see fourth step) with a factor that we found to modulate speech intelligibility in a prior proof-of-concept study (Figure S3A). This factor, which we refer to as 'ENV_H ratio' (comparable to the 'noise factor' in [70]), essentially controls the high-frequency envelope-to-noise ratio in the synthesized speech signal.

Electric stimulation

Electric stimulation parameters were set to produce relatively strong synchronous currents in the target regions, the two auditory cortices. In brief, two near-equivalent electric circuits were generated in the two cerebral hemispheres by inducing the same current in each hemisphere between a small ($5 \times 5cm^2$) stimulation electrode placed above the temporal lobe (positions T7 and T8) and a large ($10 \times 7cm^2$) common return electrode placed at the vertex (position Cz). This configuration served to center the peak of the induced intracortical current distribution on our target regions, as suggested by prior behavioral findings and electric-field simulations using a standard human head model [27, 71]. However, it could not circumvent the inherent limitation of TCS that induced currents spread widely into non-target regions, in particular the skin. As mentioned above, the shape of the applied current resembled the relevant speech rhythm.

Two-talker experiment

In the two-talker experiment, the electric stimulus was a simple f_c alternating current that resembled the prominent sinusoidal f_c envelope of the target speech signal. This current was ramped up (down) during a 10 s rest interval at the beginning (end) of each run of the experiment, respectively. EnvTCS runs involved continuous stimulation using this current, whereas in sham runs, the current was ramped down (up) during the 70 s interval that followed the initial up-ramp (preceded the final down-ramp), respectively.

Single-talker experiment

In the single-talker experiment, the electric current was shaped exactly as the low-frequency (< 16Hz) envelope portion removed from the original acoustic signal. It was obtained by summing the squared low-frequency portion of the channel envelopes across channels (see Auditory Stimulation, third step). We hypothesized that presenting the (quasi-periodic) speech envelope via envTCS modulates speech entrainment and intelligibility. More specifically, we predicted that phase spectra of envTCS and aurally-evoked residual envelope-following neural responses closely match for a specific (unknown best) audio lag that consequently strengthens entrainment and intelligibility, compared with more distant lags. For the latter non-best lags, we did not anticipate any specific pattern especially because our experimental manipulation could not induce coherent oscillatory changes (see Discussion). These predictions were based on the following considerations: First, multiplying the degraded auditory speech stimuli with the extracted broadband speech envelope substantially improved intelligibility for a specific lag, as shown by our proof-of-concept study (Figure S3A). Second, presenting auditory or visual speech-envelope information in addition to auditory speech input enhances speech-brain entrainment and intelligibility, and this benefit is disrupted by introducing lags between the envelope information and speech input

[50, 51, 65, 72–74]. Third, cortical activity entrains also to quasi-periodic electric stimulation [24] and brief intervals of TACS as short as 1800ms are sufficient to alter neural processing and perception [75, 76]. Finally, exposure to intact (envelope-carrying) speech enhances both intelligibility and temporal cortical responses associated with subsequently presented vocoded versions of the same speech [77, 78].

We fixed the maximum of the envelope-shaped current across sentences and then superimposed it onto a direct current (DC) in an amplitude ratio of 10:1, similar to conventional oscillatory DC approaches that keep the orientation of local TCS-induced currents constant within participants [79, 80]. The weak DC was applied continuously using the temporal electrodes as anodes to induce an ongoing state of enhanced neural excitability in the target brain regions [46, 71]. This was expected to increase the ability of small envelope landmarks to induce excitability changes and therewith contribute to speech entrainment. Potential transients were smoothed using a 140-ms temporal window centered on transitions between DC and envelope. Moreover, the DC was ramped up (down) during a silent 3 s rest interval at the beginning (end) of each run of the experiment, respectively.

Auditory-electric stimulus presentation

Auditory and electric stimuli were generated digitally before the experiment using a sampling rate of 16kHz and converted collectively to analog signals during the experiment using a multi-channel D/A converter (National Instruments). Stimulus timing was controlled using Datastreamer software [56]. Auditory stimuli were presented diotically via a high-fidelity soundcard (Focusrite Forte) and insert earphones (EARTone 3A). Electric stimuli were presented via two battery-operated stimulator systems (Neuroconn, Ilmenau, Germany) and rubber electrodes attached to the participant's scalp with conductive paste.

Task and experimental design

Speech intelligibility was measured using a speech recognition task requiring participants to listen to each sentence and verbally repeat as many words of it as possible. Participants responded after each sentence during a variable response interval (average duration: 5.7 s) during which their response was recorded. Participants were instructed to avoid eye movements to reduce potential visually-evoked neural phase resetting. Moreover, for the two-talker experiment, participants were instructed to focus exclusively on the male talker and ignore the female talker. Experimental conditions were defined by the audio lag, defined as the delay between the onsets of the auditory stimulus and the electric stimulus. Audio lag was varied across six equidistant steps by adjusting the inter-trial interval. The control condition was identical to the experimental conditions, except that it involved no current resembling speech rhythm. Each of the seven conditions was presented 40 times. On each of the 280 trials, a unique and novel sentence was presented. Each envTCS run involved ten repetitions of all experimental conditions. The assignment of sentences to conditions and the order in which conditions were presented within runs were individually randomized. Participants and experimenters were blinded for conditions.

Two-talker experiment

The audio lag was varied in 41.7-ms (30°) steps spanning together a whole f_c cycle. The control condition involved sham stimulation, which rendered the audio-lag manipulation virtual and left f_c neural phase unbiased. Such control trials were presented as a single block forming the sham run. Each run was composed of 60 trials and lasted approximately 9min. Four envTCS runs and one sham run were presented in individually randomized order.

Single-talker experiment

The audio lag was varied in 195-ms steps within the range from −405ms to 570ms, with positive values indicating that the auditory stimulus lagged behind the electric stimulus. This relatively large range was chosen to ensure covering the initially unknown best audio lag. The exact settings were derived from consideration of the proof-of-concept study results (Figure S3B), previous audiovisual integration results [81–84], and estimated signal transmission times [25, 85]. The control condition involved only the DC to render transcranial cues for speech entrainment unavailable to the listeners. Such control trials were randomly interleaved between experimental trials within each run. Each run was composed of 70 trials and lasted approximately 9min. Four runs were presented in individually randomized order, with half of the runs containing only sentences from the male (female) talker, respectively.

Procedure

The experimental procedure spanned two sessions involving the following steps: first, participants were seated in a sound-attenuated chamber isolated from the experimenter. Second, their hearing ability was assessed using pure-tone audiometry. Third, they were familiarized with the stimuli and task of the two-talker experiment. Fourth, they practiced the two-talker task, during which their speech recognition threshold was measured using the method of constant stimuli; this served to fix performance level across participants. The SNR was varied in five 2.5-dB steps spanning the range from −1dB to 9dB. Threshold was defined by fitting the data obtained on 55 trials with a psychometric function and identifying the SNR yielding a performance level of 45%—an intermediate level that we deemed most sensitive to the presentation of envTCS based on our proof-of-concept study (Figure S3A). For three participants, threshold was defined from the 60%-correct point (excluding these participants' data from the analyses did not alter the conclusions that can be drawn from the study). To familiarize participants with the pace of the task, a brief tone was presented shortly before the onset of each practice trial. Fifth, the last two steps were repeated for the single-talker task again without envTCS. For this, ENV_H ratio was varied in 0.5-steps spanning the range from 0.7 to 0.9 across 50 randomly ordered trials, half of which contained sentences from the male talker and the other half from the female talker. Sixth, following a break of approximately 30min, the electrodes were attached to participants' scalp, impedances were lowered to 10kΩ or less (on average 5.3kΩ), and an envTCS threshold

was estimated by adjusting peak current intensity to the point for which participants reported feeling comfortable or uncertain about the presence of the current. Seventh, stimulus parameters were set to the individually identified thresholds (SNR: 1.8 ± 2.3dB, peak current intensity: 0.9 ± 0.1mA; mean ± SD across participants) and five runs of the two-talker experiment were conducted with short breaks in between. Finally, participants were asked to provide for each run a percentage quantifying their certainty of having received electric stimulation. The second session, which took place within a few days, repeated the last three steps for the single-talker experiment. The individually identified thresholds were 0.6 ± 0.1 for ENV_H ratio and 1.0 ± 0.1mA for peak current intensity (mean ± SD across participants). Moreover, participants were asked to rate the amount of attention they paid to the electric stimulation on a four-point scale. We did not assess listeners' certainty of having received electric stimulation in the single-talker experiment because control trials occurred randomly within runs and also involved electric stimulation.

QUANTIFICATION AND STATISTICAL ANALYSIS

Data analysis

Each participant's behavioral data was analyzed as follows: First, the participant's recorded responses were scored offline by a blinded native Dutch research assistant. Second, speech performance was assessed in each condition as the percentage of correctly recognized words, computed by dividing the number of correctly recognized words by the number of presented words after pooling across trials. Third, a behavioral waveform representing (changes in) speech performance as a function of audio lag was reconstructed by concatenating the extracted behavioral measure across the six audio lags. Fourth, participants' waveforms were aligned to compensate for potential inter-individual variations in TCS-effect polarity and to therewith enable group-level analyses. Finally, effects of speech entrainment strength on speech intelligibility were identified by statistically testing the aligned waveforms for systematic changes across lags.

Regarding the fourth step, the polarity of TCS effects depends on various factors including the prepolarization (baseline excitability) of the task-relevant neuronal population [86] and the orientation of this population relative to local TCS-induced currents [25, 87]. Because folding patterns and speech-entrainment loci in auditory cortex (our target region) tend to vary greatly across individuals [28, 88], they together could potentially produce opposite relative neuron-current orientations among some of our participants (e.g., on opposite walls of a gyrus). Our approach for compensating for reversed (excitatory versus inhibitory) TCS-effect polarities presumes that opposite behavioral patterns reflect reversed TCS-effect polarities. We did not position TCS electrodes based on prior individualized, functional-anatomical neuroimaging-informed current-flow simulations, which is an alternative, more labor-intensive, and less commonly applied approach [45].

Two-talker experiment

Initial exploration of individual performance waveforms revealed that the distribution of the audio lag for which participants performed best did not deviate significantly from uniformity (Figure 2A). This observation of a highly variable 'best lag' across participants hints at inter-individual variations in TCS-effect polarity (see previous section). To compensate for such variations, we associated each participant's maximum entrainment strength with the participant's best lag and phase-wrapped the remainder of the behavioral waveform under the hypothesis that envTCS influenced behavior via neural entrainment. Under this hypothesis, the aligned waveform should exhibit an f_c cycle. More specifically, lags near the best lag (distance from best lag: −60°–60°) should delimit an excitatory f_c half-cycle associated with relatively good behavioral performance, whereas more distant lags (distance from best lag: 120°–240°) should delimit the opposite, i.e., an inhibitory half-cycle associated with poorer performance. To test this key prediction—and thereby verify our hypothesis—performance was averaged across the presumed excitatory half-cycle (best-lag distances −60° and 60°) and the presumed inhibitory half-cycle (best-lag distances 120° and 240°), and the two resulting averages were compared statistically. The aligned best lag (0°) and its counterphase lag (180°) were excluded from this analysis to avoid circular reasoning and unbalanced samples, respectively. Including each or both of these lags in this analysis did not alter the conclusions that can be drawn from the study.

Subsequently, we assessed whether behavior fluctuated primarily at f_c to verify that the observed fluctuation indeed reflected envTCS-induced neural entrainment. For this, individual spectral densities were computed from the individual performance waveforms using the discrete Fourier transform and the magnitudes of the resulting frequency bins were compared. All data points could be included in this spectral analysis as the magnitude spectrum is generally unaffected by phase shifts as those induced by the best-lag alignment. Given the limited number of data points and sampling rate (six phase bins spanning one f_c period) only three bins centered on $1f_c$, $2f_c$, and $3f_c$ could be resolved.

The control analysis involved data obtained during sham stimulation. These data were stratified according to the 'virtual' audio lag, i.e., the experimental condition that would have occurred if the alternating current had been left on. Data acquired during the on/off ramps were not considered.

Single-talker experiment

Sentences were first temporally segmented into word intervals to enable focusing the analysis on those intervals during which envTCS was presented. Potentially confounding effects of interval were excluded by subtracting the control condition from envTCS conditions separately for each interval. The resulting behavioral measure quantifies the listener's benefit from envTCS in units of percentage points (pp), independent of performance level.

Initial exploration of individual speech-benefit waveforms revealed that the distribution of the audio lag for which participants benefitted maximally did not deviate significantly from uniformity ($D_{22,100}$ = 0.25, p = 0.18), which hints at potential inter-individual

variations in TCS effect-polarity, as in the two-talker experiment. However, an alignment based on participants' best lag as above was not applicable here because stimuli had no fixed, strictly cyclical structure and local current orientation was fixed within participants due to the constant DC offset. Instead, an approach based on participants' overall TCS-effect polarity was applied presuming that envTCS at the (unknown) best lag induces a positive benefit; thus we interpreted overall negative envTCS benefits (i.e., no positive benefit at any lag) to arise from overall reversed (inhibitory) TCS effects. Indeed, for seven participants, such an overall negative benefit was observed. Moreover, these participants' most and least beneficial lags ('best' and 'worst' lag, respectively) appeared to be polarity-reversed: although this participants' *best*-lag distribution was relatively flat, its maxima included the *worst* lag most frequently observed among the remaining 15 participants, and vice versa for the worst-lag distribution (Figure S4A versus S4B). To compensate for these inter-individual variations, we aligned these seven participants' behavioral waveforms to the other participants' waveforms by inverting their sign. Alternatively excluding these data from the analyses did not alter the conclusions that can be drawn from the study. The resulting best-lag distribution of the group was found to concentrate exclusively on 375ms and deviate significantly from uniformity (Figure 4A).

As the applied polarity-based alignment approach presumes a positive envTCS benefit for an undefined lag, it potentially induces an overall, positive bias in group-level statistics of benefit; therefore we considered only between-condition differences, not absolute values, of this measure.

Statistical analysis

Participants' individual measures were submitted to second-level (random-effects) group analyses using parametric statistical tests (ANOVA and paired t test). Assumptions of normality and sphericity were verified with Kolmogorov-Smirnov tests and Mauchly's tests respectively, which did not detect any significant deviation from normality or sphericity. Best-lag distributions were tested for non-uniformity with a Rayleigh *z*-test (two-talker experiment) and two-sample Kolmogorov-Smirnov tests based on the *D*-statistic (single-talker experiment). A significance criterion $\alpha = 0.05$ was used and type-I error probabilities inflated by multiple comparisons were corrected by controlling the false-discovery rate [89]. Effect sizes were quantified using eta-squared (η^2) or Cohen's *d*. Reported summary statistics represent mean ± SEM across all participants unless stated otherwise.

DATA AND SOFTWARE AVAILABILITY

Data are available in Mendeley Data at https://doi.org/10.17632/zwd67kjpdd.1.

Current Biology, Volume *28*

Supplemental Information

Neural Entrainment to Speech Modulates Speech Intelligibility

Lars Riecke, Elia Formisano, Bettina Sorger, Deniz Başkent, and Etienne Gaudrain

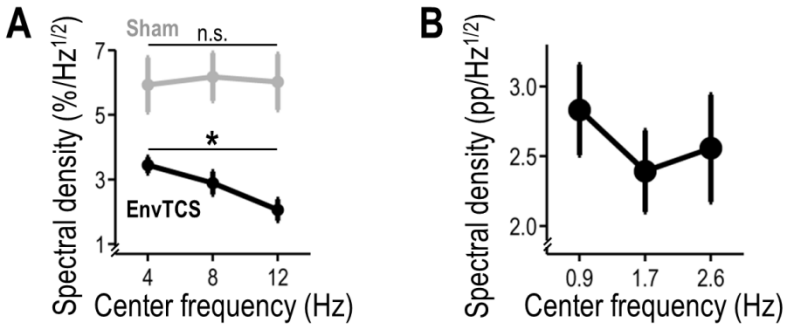

Figure S1. Spectral-analysis results. Related to Figures 2C and 4C.
A. Spectral results from the two-talker experiment, related to Figure 2C. To enable assessing potential contributions from oscillatory neural excitability fluctuations beyond the critical delta/theta range, this plot shows the average magnitude spectrum (mean±s.e.m. across listeners) of individual performance waveforms (see exemplary black waveforms in Figure 2B). This spectrum illustrates the size of behavioral modulations induced by best-lag distance as a function of the frequency of the presumed underlying neural oscillation. During envTCS (black), spectral magnitude peaked at the frequency bin centered on the envTCS frequency (f_c=4Hz) and decreased monotonically across higher frequencies; thus the observed behavioral modulations were resembled better by oscillations closer to the envTCS frequency. This observation was supported by a one-way ANOVA including oscillation frequency (4Hz, 8Hz, 12Hz) as factor, which revealed a main effect on spectral magnitude ($F_{2,40}$=4.45, η^2=0.14, corrected P=0.036). Post hoc tests showed that the 4-Hz oscillation was significantly more explanatory than the 12-Hz oscillation (t_{20}=3.61, d=0.79, corrected P=0.0036) but not the 8Hz oscillation (t_{20}=1.13, corrected P=0.19). These results show that the observed cyclic effect of best-lag distance on speech performance is frequency-selective, underscoring that it stems from entrainment of neural oscillations with frequencies close to the speech rhythm (in the delta/theta range).
Control analyses based on the virtual-lag sham data (gray) did not replicate the observed spectral peak at 4Hz or the oscillation-frequency effect on spectral magnitude ($F_{2,40}$=0.03, corrected P=0.97). A two-way ANOVA on spectral magnitude, including stimulation condition (envTCS vs. sham) and oscillation frequency (4Hz, 8Hz, 12Hz) as factors, yielded no significant interaction, suggesting that although significant frequency selectivity was observed exclusively for the envTCS data, this selectivity was only slightly stronger than for the sham data. * corrected P<0.05, n.s. non-significant.
B. Spectral data from the single-talker experiment, related to Figure 4C. Analogously to the data from the two-talker experiment in panel A, this plot shows the average magnitude spectrum (mean±s.e.m. across listeners) of individual speech-benefit waveforms (e.g., see waveforms in Figure 4B) to illustrate the size of audio lag-induced changes in speech benefit as a function of the frequency of potentially underlying neural oscillations. Benefit changes peaked at the frequency bin centered at 0.9Hz and varied non-significantly across higher frequency bins. Only summary statistics are reported here because stimulation was not fixed or strictly oscillatory in the single-talker experiment and we had no hypothesis regarding cycles in speech benefit.

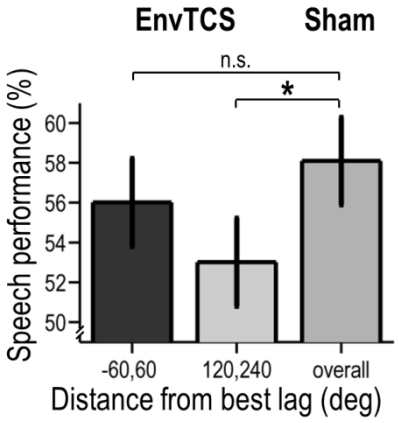

Figure S2. EnvTCS induced impairment in the two-talker experiment. Related to Figure 2D.

To assess potential benefits of envTCS for intelligibility in the two-talker experiment, we compared speech performance during each presumed envTCS-induced half-cycle vs. sham stimulation. The bar on the right shows overall performance (mean±s.e.m. across listeners) during sham stimulation (i.e., without stratification for virtual lag) and the other bars are the same as in Figure 2D left. Statistical analysis revealed a significant suppressive effect of envTCS during the presumed inhibitory half-cycle (on average -5.1±2.0 percentage points, t_{20}=-2.53, d=-0.55, corrected P=0.03), but no benefit (excitatory half-cycle: t_{20}=-1.29, corrected P=0.89; averaged across all inhibitory and excitatory lags: t_{20}=-1.24, corrected P=0.34). We did not analyze the overall benefit of envTCS in the single-talker experiment, because its strength could be biased due to the polarity alignment. In sum, the observation of a significant impairment in the two-talker experiment indicates that envTCS, if applied at 'appropriate' latency, can hamper the intelligibility of a talker in the listener's focus of attention. * corrected P<0.05, n.s. non-significant.

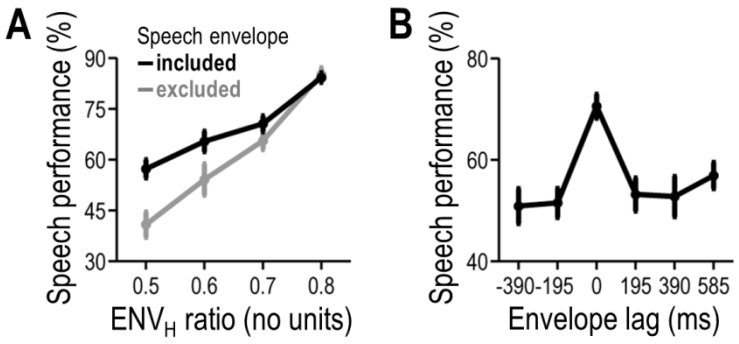

Figure S3. Results from proof-of-concept study for the single-talker experiment. Related to STAR Methods.

Data shown in this figure were collected from another fifteen normally-hearing listeners performing the single-talker task without envTCS.

A. Speech performance (mean±s.e.m. across listeners) is plotted as a function of ENV_H ratio, a parameter controlling the high-frequency envelope-to-noise ratio in the synthesized speech signal. Auditory speech stimuli were identical to those in the main experiment (gray), or further multiplied with the wideband speech envelope that defined the envTCS in the main experiment (black). The latter condition served to simulate an aural, not transcranial, application of envTCS. A two-way ANOVA including ENV_H ratio and envelope presence as factors revealed main effects on speech intelligibility (ENV_H ratio: $F_{3,42}$=81.9, η^2=0.62, corrected P=10^{-7}; envelope presence: $F_{1,42}$=21.0, η^2=0.06, corrected P=0.0006) and a significant interaction ($F_{3,42}$=3.93, η^2=0.04, corrected P=0.015). These results show that (i) ENV_H ratio can be utilized to experimentally manipulate the intelligibility of the auditory speech stimuli used in the main experiment and (ii) aural presentation of speech-envelope information improves the intelligibility of these stimuli, especially at low performance levels.

B. Speech performance (mean±s.e.m. across listeners) is plotted as a function of envelope lag. This parameter, which was analogous to the audio-lag parameter in the main experiment, controlled the delay between the auditory speech stimulus as presented in the main experiment and the speech envelope with which that stimulus was convolved here. ENV_H ratio was fixed to 0.7. A one-way ANOVA including envelope lag as factor revealed a main effect on speech intelligibility ($F_{5,70}$=9.68, η^2=0.25, P=10^{-7}). This result shows that the benefit from aurally presented speech envelope (panel A) depends on the relative timing of this envelope.

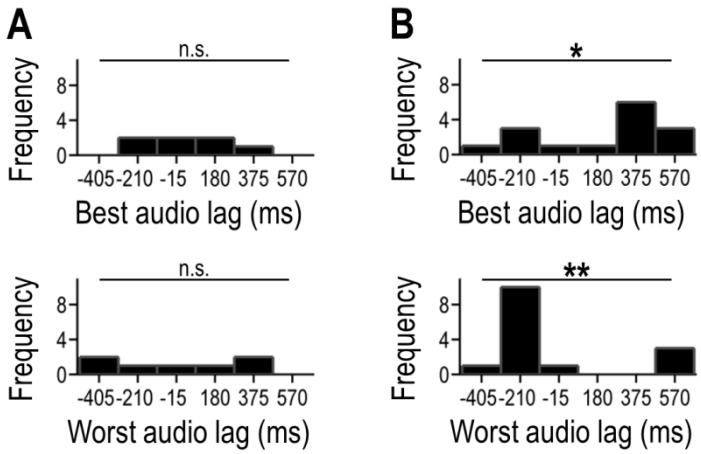

Figure S4. Initial data exploration for the single-talker experiment. Related to Figure 4A and STAR Methods.

A. Best-lag distribution (top) and worst-lag distribution (bottom) for participants in the single-talker experiment who showed exclusively negative envTCS benefits (i.e., no positive benefit at any lag) before alignment. These participants' distributions did not deviate significantly from uniformity ($D_{7,100}$=0.36, corrected P=0.40 and $D_{7,100}$=0.29, corrected P=0.58, respectively).

B. Same as panel A, but for all other participants in the single-talker experiment. These participants' best-lag distribution and worst-lag distribution deviated significantly from uniformity ($D_{15,100}$=0.41, corrected P=0.034 and $D_{15,100}$=0.53, corrected P=0.0025), revealing a peak at the 375-ms lag and -210-ms lag, respectively. These peaks fell respectively among the worst lags and best lags of most participants shown in panel A, suggesting that the two groups differed in TCS-effect polarity. These putative polarity differences were compensated in the analysis by inverting the sign of the behavioral waveforms of the participants shown in panel A.

*, ** corrected P<0.05, 0.005, n.s. non-significant.