Brain Stimulation 13 (2020) 1402–1411

Contents lists available at ScienceDirect

Brain Stimulation

journal homepage: http://www.journals.elsevier.com/brain-stimulation

Pulsed transcranial electric brain stimulation enhances speech comprehension

Philipp Ruhnau [a, b, *], Katharina S. Rufener [a, b], Hans-Jochen Heinze [a, b], Tino Zaehle [a, b]

[a] Department of Neurology, Otto-von-Guericke-University, Magdeburg, Germany
[b] Center for Behavioral Brain Sciences, Otto-von-Guericke-University Magdeburg, Germany

ARTICLE INFO

Article history:
Received 23 October 2019
Received in revised form
27 May 2020
Accepted 21 July 2020
Available online 28 July 2020

Keywords:
Entrainment
Envelope
Speech processing
tES

ABSTRACT

Background: One key mechanism thought to underlie speech processing is the alignment of cortical brain rhythms to the acoustic input, a mechanism termed entrainment. Recent work showed that transcranial electrical stimulation (tES) in speech relevant frequencies or adapted to the speech envelope can in fact enhance speech processing. However, it is unclear whether an oscillatory tES is necessary, or if transients in the stimulation (e.g., peaks in the tES signal) at relevant times are sufficient.
Objective: In this study we used a novel pulsed-tES-protocol and tested behaviorally if a transiently pulsed - instead of a persistently oscillating - tES signal, can improve speech processing.
Methods: While subjects listened to spoken sentences embedded in noise, brief electric direct current pulses aligned to speech transients (syllable onsets) were applied to auditory cortex regions to modulate comprehension. Additionally, we modulated the temporal delay between tES-pulses and speech transients to test for periodic modulations of behavior, indicative of entrainment by tES.
Results: Speech comprehension was improved when tES-pulses were applied with a delay of 100 ms in respect to the speech transients. Contradictory to previous reports we find no periodic modulation of behavior. However, we find indications that periodic modulations can be spurious results of sampling behavioral data too coarsely.
Conclusions: Subject's speech comprehension benefits from pulsed-tES, yet behavior is not modulated periodically. Thus, pulsed-tES can aid cortical entrainment to speech input, which is especially relevant in a noisy environment. Yet, pulsed-tES does not seem to entrain brain oscillations by itself.

Introduction

To encode language properly our auditory system needs to efficiently extract linguistically meaningful information from the acoustic input. Slow modulations within the speech stream conveyed by the amplitude envelope represent information at the syllabic scale. Accordingly, the auditory system needs to identify these temporal features in the acoustic input to segregate the speech stream. In short, timing in speech processing is essential [1] and therefore coding temporal cues (transients such as syllables) in the acoustic input will determine the speech comprehension success.

Previous work showed that spoken language is reflected in phase aligned brain activity in a hierarchical network [2]. This process, typically in the frequency range of cortical theta and matching syllable rates, is termed speech-brain entrainment [3–5]. It is assumed that the quasi-rhythmicity of speech leads to entrainment of brain oscillations.

Transcranial electric stimulation (tES) based on the shape of the speech amplitude envelope has been demonstrated to enhance processing of sentences presented in noise [6] and comprehension of speech material where the envelope has been degraded [7]. These studies aimed to confirm the relevance of entrainment to low level speech features [5] such as the amplitude envelope for speech comprehension. More specifically, these studies [6,7] extracted the speech amplitude envelope and applied tES mimicking the extracted signal (thus called envelope tES) to auditory cortex regions. Importantly, envelope tES was applied at different temporal delays together with a corresponding speech signal to test for periodic modulations of behavior. Both studies found phasic

* Corresponding author. Otto-von-Guericke University, Department of Neurology, Leipziger Str. 44, 39120, Magdeburg, Germany.
E-mail address: mail@philipp-ruhnau.de (P. Ruhnau).

https://doi.org/10.1016/j.brs.2020.07.011
1935-861X/

modulations of speech comprehension based on the speech to tES delay, interpreted as indicative of entrainment by envelope tES.

Here we test whether a stimulation protocol that is based on transients in the speech signal - but without the continuous oscillatory modulation from previous envelope tES studies - can still lead to behavioral benefits. TES was applied transcranially with targets in bilateral auditory cortex regions. We used a pulsed-tES protocol aiming to phase align cortical activity around meaningful transients in speech (i.e., syllables) to enhance speech comprehension.

In addition, we modulated the temporal delay of the tES-pulses in respect to the presentation of the speech signal to test for a phasic modulation of behavior, an indicator of entrainment caused by tES. We hypothesize that tES applied with a critical delay of around 100 ms will be most beneficial to speech comprehension in line with previous results [6].

Typically, studies testing entrainment sample the (most of the time behavioral) effects of interest with only few data points, resulting in temporal sampling that show the highest sensitivity for the frequencies of interest. For instance, when investigating a rhythm relevant for speech processing, cortical theta (~4Hz, one cycle 250 ms), five to six data points are used regularly (e.g., Refs. [6,8]). However, this could create a bias towards hypothesized behavioral modulations [9]. We used more samples (higher temporal sampling) to estimate this bias in our own observed data as well as using computer simulations. We hypothesize that coarse temporal sampling (few measured data points) can bias results in favor of entrainment, a bias that disappears at increasing sampling rates.

Materials and methods

Subjects

For the current study we recruited twenty-four healthy young adults, this sample size was chosen based on the range of previous envelope-tES/tACS studies [6,7,10]. All but two subjects, according to their own report, were right handed and none reported a history or current neurological or psychiatric disease. All were of normal hearing. Three subjects had to be excluded because the current stimulator did not apply any current due to a technical malfunction, another subject received the wrong stimulation setup (different electrode size), thus the final sample consisted of 20 subjects (age M = 21.9, SD = 2.25, range 19–30 years, 11 female). The study was approved by the ethics committee of the medical faculty of the Otto von Guericke University Magdeburg and carried out in accordance with the Declaration of Helsinki. All subjects gave written and oral consent prior to the study.

Sentence comprehension test

Following a study that was one of the first to use envelope-tES [6] we used the Oldenburg sentence test (Oldenburger Satz Test - OLSa [11]) and followed its manual. In brief, the OLSa is an adaptive speech comprehension in noise test, which consists of 100 semantically unpredictable, unique sentences embedded in noise, frequency-matched to the speech signal. The test consists of 40 lists of 30 sentences each, each of the 100 sentences appear in multiple lists but are never repeated within one list. Every sentence has the same 5-word structure (name, verb, number, adjective, object). The subjects' task is to listen to the sentences and repeat as many words as they can. The test measures the sentence comprehension threshold (SCT) in dB Signal/Noise difference such that the noise is presented at a constant level (here 68 dB SPL) and the intensity of the sentences (signal) is adjusted adaptively depending on the subjects' performance. Typical performance of normal hearing subjects is around −7 dB S/N [12], that is sentences can be presented at 7 dB below the intensity of the noise for the subjects to understand about half of the words (2–3 items). The sentence material of the OLSa is designed in such a way (arbitrary assignment of words to a sentence so it might appear nonsensical) that it is not possible to memorize the content and therefore the test can be repeated and still produce similar results (see OLSa manual as well as [11]).

Pulsed transcranial electric stimulation (pulsed-tES)

Pulsed-tES was applied using three rubber electrodes with a DC Stimulator Plus (Neuroconn, Germany). Two 5 × 5 cm electrodes, used as linked anodes, were placed over T7 and T8 to target auditory cortex regions and a larger 5 × 10 cm electrode, used as cathode/return, was placed on Cz. This setup was identified to stimulate auditory regions using modeling [13] and used in previous studies investigating language processing [6–8]. Electrodes were fixed on the scalp with a conductive paste (Ten20, D.O. Weaver, Aurora, CO, USA) and impedances were kept below 10 kOhm if possible.

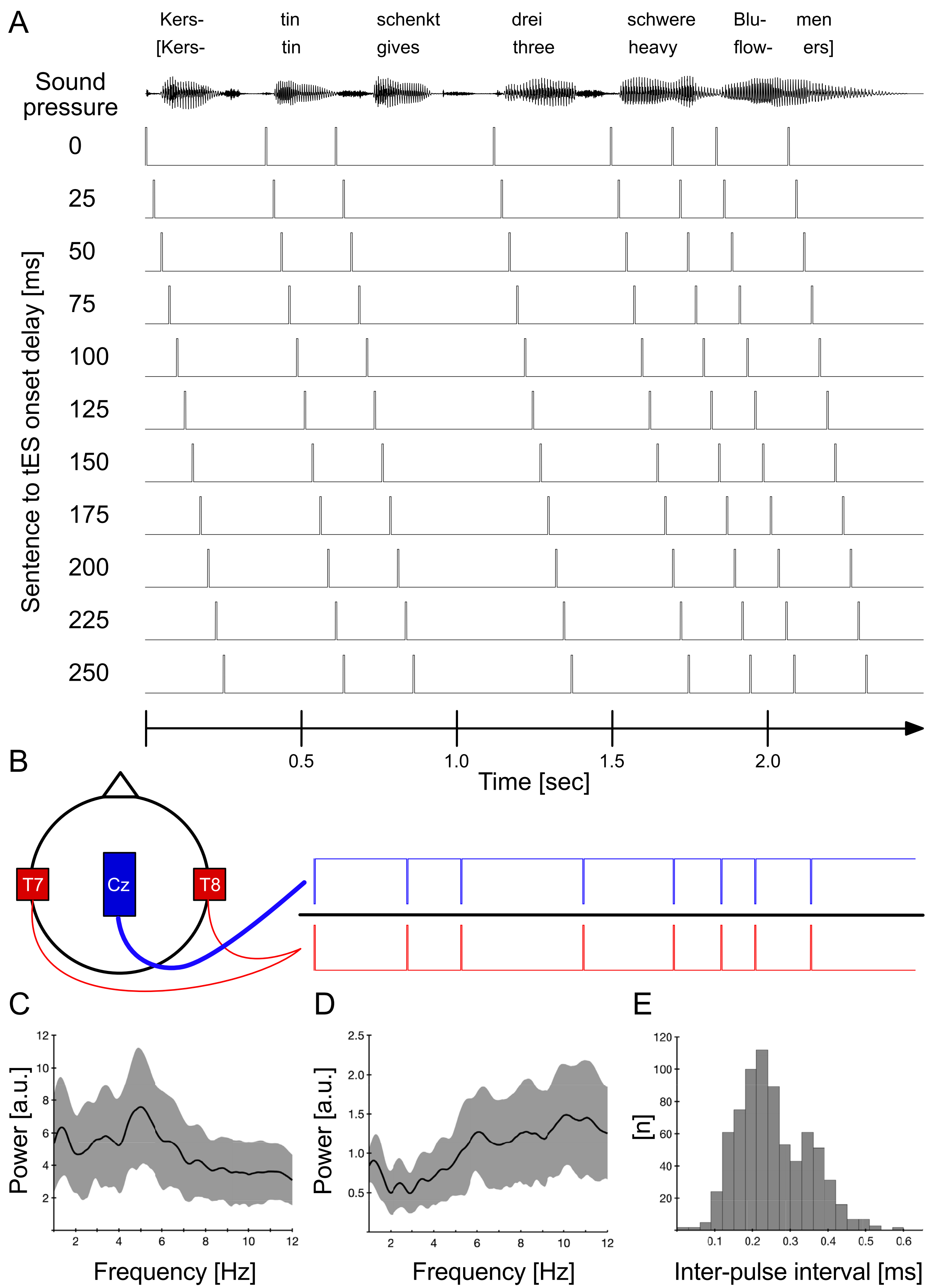
The applied current consisted of brief (5 ms) DC pulses at time points between syllables, identified in the spectrogram as points of lowest energy in the amplitude-envelope (see Fig. 1). Then, for the 11 experimental blocks the latency between the pulsed-tES and the sentence presentation onset was modulated from 0 to 250 ms in 25 ms steps (Fig. 1). The order of the 11 latency blocks was randomized across participants. Studies using tACS that aim to investigate phasic modulation of perception typically use lower sample rates as their main interest is in slow frequencies (e.g., Refs. [6,7,10], for sampling rates of 20, 23, and 25 Hz, respectively). However, coarse sampling rates could have biased previous results. For instance, the resolvable spectrum will be limited to the Nyquist frequency (sampling rate divided by 2) and below, so potential faster rhythms cannot be recovered.

Synchronous auditory-electric stimulation

Auditory and pulsed-tES signals were created digitally using Matlab (Mathworks, NA, USA) at a sampling rate of 44100 Hz. They were then converted to analog signals in a D/A converter (NI-6212, National Instruments), which was connected to the DC stimulator linked to the rubber electrodes (see above) and an amplifier (Sony DTC-57ES) driving headphones (Sennheiser, HD 65 TV). The timing of the two signals was controlled with a digital oscilloscope (Red Pitaya, Solkan, Slovenia).

Experimental design and general procedure

Subjects sat comfortably in a dimly lit room. After stimulation electrodes attachment (see pulsed-tES procedure above) the tES sensation threshold was estimated using two 1-up/1-down staircases [14], one starting at 0 mA and one at 0.8 mA. The pulsed-tES sequences were selected randomly from the sample used in the experiment. After each of the subjects' responses the tES intensity was increased ('no' response) or decreased ('yes' response). After response reversals the increment/decrement steps were adjusted adaptively (0.2, 0.1, 0.05, 0.025 mA). The step size was not adjusted for the first two reversals. After a total reversal of 12 or a total of 30 trials per staircase the procedure was stopped and the threshold estimated as the average over the last n-2 reversals. To ensure that the staircase procedure worked as expected (two intensities converging to a similar level) the intensities over trials were plotted, which the experimenter examined and decided whether to


A
Kers-
tin
schenkt
drei
schwere
Blu-
men
[Kers-
tin
gives
three
heavy
flow-
ers]
Sound pressure
Sentence to tES onset delay [ms]
0
25
50
75
100
125
150
175
200
225
250
0.5
1.0
1.5
2.0
Time [sec]
B
T7
Cz
T8
C
Power [a.u.]
Frequency [Hz]
D
Power [a.u.]
Frequency [Hz]
E
[n]
Inter-pulse interval [ms]

repeat the staircase procedure or whether to continue. The tES intensity for each subject was set to 0.05 mA below the estimated threshold. The maximally applied current was 1.95 mA (2 mA maximal value of the staircase minus 0.05 mA).

Subsequently, the subjects did two training blocks of the sentence test (see above) after which 11 blocks were run, randomly assigned to a tES-sentence delay (see Fig. 1). The net experimental time was about 55 min. With preparation, breaks and post-experimental procedure (removal of the electrodes, debriefing, hair washing) the experiment did not exceed 2.5h.

After each block the subjects indicated whether they felt the electric stimulation or not. The experiment was double-blind, that is, neither experimenter nor subject knew the order of conditions while the experiment was run.

Descriptive analysis of sentence envelopes and pulsed tES

To descriptively evaluate the rhythmicity, both, sentence amplitude envelopes as well as pulsed tES sequences, were analyzed in regard to their spectral amplitude using a fast Fourier transform (Matlab's fft function). Sentence waveforms were normalized to 1 and their amplitude envelopes were extracted using Hilbert transforms. Both envelopes as well as pulsed tES sequences were detrended before the Fourier transform.

Additionally, we computed a histogram of the inter-pulse intervals (i.e., the temporal difference between adjacent tES pulses in a sequence) to furthermore evaluate temporal regularities in the tES trains.

Statistical analysis of performance distribution and peak performance

We tested whether SCTs were equally distributed across all latency bins using a Kolmogoroff-Smirnoff test. Given previous results [6] we expected strongest benefits from pulsed-tES at 100 ms delay from sentence onset. Based on the distribution test and previous results we tested the SCT at the 100 ms bin against the average of all other bins using a paired *t*-test. Note that, in contrast to a best performance vs. sham comparison that is biased to find a positive effect even in random data [9], our analysis is not. This was ensured by using Monte Carlo simulation (n = 5000 random data sets with similar mean (M), standard deviation (SD) and n as our data), in each data set a *t*-test was performed (100 ms bin vs. average of other bins) resulting in a simulated t-distribution with M = 0.013, SD = 1.08 (normal distribution).

Statistical fitting procedure

If not stated otherwise least squares fit using the *lsqcurvefit* function (in Matlab's Optimization toolbox) was used, which by default allows for 1000 iterations to converge on the best parameters. Analogue to previous approaches [6,15] we fitted a sine wave to the subjects SCTs using the formula $y = a + b*sin(c*2*\Pi*x + d)$, where x is the pulsed-tES to sentence time lags from 0 to 250 ms. a, b, c, and d were the parameters estimated by the fitting procedure. Parameter a was the intercept ([lower bound, upper bound]; [-4 4]), b the amplitude [-2 2], c the frequency [2 20], and d the phase delay of the sine wave [-2*π 2*π]. We restricted the parameter space similar to Ref. [6], note however, that our frequency range was larger because of higher sampling rate and thus nyquist frequency. Furthermore, a non-restricted fitting procedure led to similar estimated parameters and BIC values as the restricted fits.

The quadratic fit was estimated using the formula $y = a*x^2 + b*x + c$, where x again is the pulsed-tES to sentence delays and the parameters are a for the quadratic coefficient, b for the linear coefficient and c for the intercept.

The linear fit was estimated using the formula $y = a*x + b$, with a as the linear coefficient and b as the intercept.

Model fits (sine wave, quadratic, linear) were compared using the Bayesian information criterion [16], a goodness-of-fit parameter based on the residuals using the following formula:

$$BIC = n + n*ln\frac{\sum R^2}{n} + \ln(n)*p$$

where R is the residuals, n is the number of trials, and p the number of parameters in the model. BICs were computed for every fit per participant and then averaged, the results were compared descriptively, with lower BIC values indicating a better fit. The BIC penalizes fits with many parameters and thereby corrects for the (higher parameter n based) advantage of the sine wave fit over the other two.

Comparison of different sampling rates

Given that we used a higher sampling rate (40Hz) to sample behavior (as opposed to e.g. Ref. [6,7,10], for sampling rates of 20, 23, and 25 Hz, respectively) we aimed to compare fitting procedures for our data at different (40 Hz; 20 Hz) sampling rates. To do so we repeated the fitting procedure described above on a reduced dataset by omitting every second data point (resulting in tES delays from 0 to 250 ms in 50 ms steps, similar to Ref. [6]).

To test for biases in the testing of lower sampling rates, we furthermore evaluated the fitting procedure on random numbers (n, mean, and standard deviation set similar to our data) with a low (20 Hz) and a higher (40 Hz) sampling rate. We used Monte Carlo simulations [17] (n = 5000) to create group level BIC distributions (see above) to compare different fitting procedures (linear, quadratic, sinusoidal).

Results

Sub-threshold tES parameters and sensation

Following the tES threshold estimation the mean stimulation intensity was M = 0.73 mA with an SD = 0.59 (see Table 1). Subjects reported to have felt the tES on average in 1.7 out of 11 blocks (SD = 0.63 blocks, range 0–8). Since the stimulation was present in all blocks we can assume most subjects (all but two who felt tES in more than half of the blocks) were sufficiently blind to the experimental condition. Analyses below were run on all subjects but repeated on the subset excluding the two subjects who felt the stimulation, however, results did not change.

Fig. 1. Study design and material. (A) Sentences were presented with a pulsed transcranial electric stimulation (pulsed-tES) signal at varying onset latencies (top). Electric pulses (5 ms) were presented at time points between syllables (lowest amplitude envelope point before each syllable). (B) Positive pulses where applied over auditory cortex regions using two 5 × 5 cm electrodes over T7/T8, and a negative pulse was applied at a larger electrode at Cz (5 × 10 cm, 'return' electrode). (C) Average power spectrum of all 100 sentence amplitude envelopes showing a peak around 5 Hz, confirming the quasi-rhythmicity. (D) Average power spectrum of all 100 tES sequences derived from the sentences. No clear peak emerged. Shaded areas represent the standard deviation. (E) Inter-pulse interval histogram showing a peak around 200 ms (corresponding to the syllable rate at 5Hz).

Table 1
Stimulation parameter per subject.

Subject #	tES intensity [mA]	Impedance [kΩ]	tES sensation reported [n blocks of 11]
1	0.28	5	0
2	0.5	10	0
3	0.3	9	6
4	0.22	18	0
5	0.51	10	2
6	0.16	7	2
7	0.55	7	0
8	0.72	10	0
9	1.29	2	4
10	1.20	14	1
11	0.23	6	3
12	0.2	7	2
13	0.19	3	1
14	1.75	6	0
15	1.25	3	0
16	0.11	6	1
17	1.56	1	1
18	1.95	2	8
19	0.76	2	1
20	0.9	1	0

Effects of tES delay

Sentence comprehension thresholds (SCTs) where on average at 6.86 dB S/N, which is in the expected range of normal hearing populations [12].

SCTs (see Table 2) where not uniformly distributed but showed a concentration at a preferred delay bin (Kolmogorov-Smirnov test, $d = 0.64$, $p = 0.012$). The 100 ms delay bin showed most frequently the smallest (i.e., best) SCT value (Fig. 2B). In the grand-average (Fig. 2A), the 100 ms delay also showed the best performance compared to the average of all other bins ($t(19) = 1.95$, $p = 0.034$, one sided). Given that the grand average showed the local minimum at 100 ms and we are thus comparing the minimum of our distribution with the rest of the distribution our analysis could be affected by a bias similar to maximum vs. minimum testing (cf. Ref. [9]). Therefore, we ran a permutation test on our experimental data. In $n = 10000$ permutations we shuffled the tES-to-sentence bins within each subject. We then computed the difference between the minimum bin (lowest SCT in dB Signal/Noise difference, thus, best performance) determined from the grand-average (average over all subjects) and the mean of the remaining bins in each permutation. This difference distribution should capture the bias if selecting the maximum from the grand-average post-hoc. As expected, there is a bias in selecting the peak post-hoc (evident by the negative mean of the distribution, $M = -0.195$, Fig. 2C), yet our observed value (-0.278) is at $p = 0.0676$, and thus smaller than >93% of the distribution, increasing the confidence in this finding.

The worst performance was most often recorded at the 0 ms tES delay (not shown).

Sine fitting results

Descriptively, and contrary to previous findings [6] our data do not show sinusoidal modulations on the individual subject level (Fig. 3). This is confirmed by the BIC analysis. When comparing the three fitted models (linear, quadratic, sine) the linear fit yielded the lowest BICs, sine and quadratic fit yield larger BICs and are of similar magnitude. However, BIC differences of around 2 or smaller are 'not worth more than a bare mentioning' [18,19], and thus no model turns out as a clear winner.

We hypothesized that previous work might have found sinusoidal fits because of the low sampling rate of the behavioral measure (e.g. Ref. [6,7], used sampling rates of 20 and 23Hz, respectively). To test this, we created a comparable dataset by omitting every second data point in our data and ran the fitting procedure again. In fact, using this reduced samples data set we replicated previous results [6]. Sinusoidal patterns emerge very clearly from the behavioral SCT patterns (Fig. 4A and B) and were confirmed in the BIC analysis (sine fit < linear and quadratic fit; Fig. 4C). The estimated average parameters of the winning sine fit are $y = -1.2113 + 0.0939 * \sin(4.3690*2\pi*x + 0.0147)$, the estimated frequency of the sine (4.4Hz) is perfectly in the range of neuronal theta oscillations, as was previously found.

Monte Carlo simulation results for random data and fitting bias

To follow up on the fitting results we ran Monte Carlo simulations to test for a bias in the data with the lower sampling rate [17]. When feeding 6 randomly created data points (=20Hz sampling rate, Fig. 5A) into the fitting procedure the sine fit has an advantage to fit random data better than the linear and quadratic fit (sine median (Mdn) = −6.03, 95% confidence interval (CI) = [-9.77 -2.79]; linear Mdn = −1.48, CI = [-3.46 0.17]; quadratic Mdn = −1.97, CI = [-4.47 0.04]; Fig. 5A), whereas when using double the data points (40Hz sampling rate, Fig. 5B) the linear fit has a numerical advantage over the other two models but the confidence intervals overlap with the medians (sine Mdn = −0.15, CI = [-2.72 2.16]; linear Mdn = −0.97, CI = [-2.85 0.77]; quadratic Mdn = −0.02, CI = [-2.09 1.89]; Fig. 5B). Thus, while the low sampling rate data clearly suffer from a bias that favors a sinusoidal

Table 2
Average sentence comprehension thresholds (SCTs) and standard deviation in the different delay conditions.

tES-sentence delay [ms]	0	25	50	75	100	125	150	175	200	225	250
Mean SCT [dB]	−6.80	−6.72	−6.88	−6.86	−7.11	−6.95	−6.85	−6.92	−6.73	−6.92	−6.83
SCT standard deviation [dB]	0.72	0.50	0.92	0.83	0.80	0.69	0.67	0.84	0.76	0.78	0.82

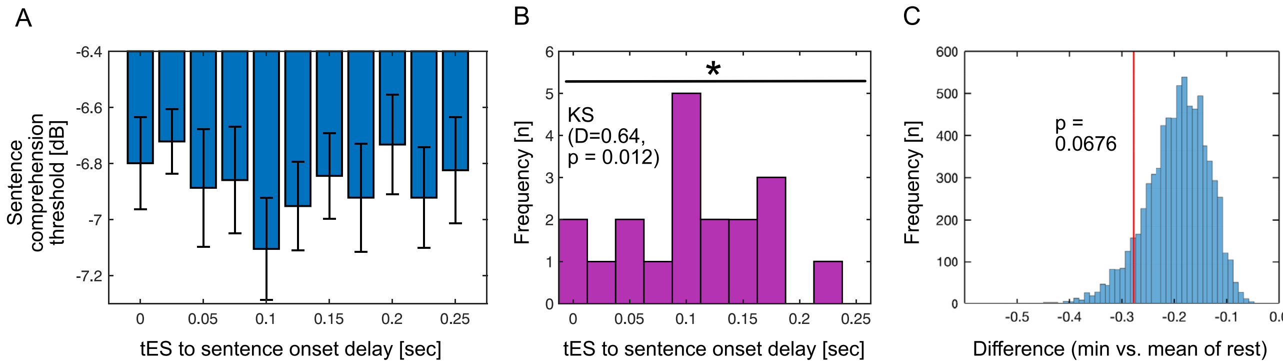


Fig. 2. Sentence comprehension performance by tES delay conditions. (A) SCT for the individual tES delays with maximum performance at 100 ms tES delay. Error bars represent standard error of the mean. (B) Histograms showing the tES delay bin that provided the minimum sentence comprehension threshold (SCT) in each subject, i.e., their best performance. The distribution is non-uniform (Kolmogorov-Smirnov). (C) Permutation test result ($n = 10{,}000$), controlling for a bias that can be introduced by post-hoc selection of the grand average minimum. In each permutation, all subjects SCTs were shuffled across delay bins and the grand average over subjects computed. Then, from that grand average the minimum was selected and subtracted from the average of the remaining SCTs. The differences from all permutations form the here shown distribution. The observed value from our experiment is at 6.76% of the distribution.

fit as proven with the Monte Carlo random data simulation, the 40 Hz sampling rate data does not.

Based on these results we repeated this procedure for a range of sampling rates starting at 16 Hz up to 61 Hz in 5 Hz steps (1000 Monte Carlo simulations per frequency) to evaluate when a bias caused by the sampling rate emerges and when it disappears. For each simulation we computed the difference between all pairs of the three model fit BIC results. These simulations (Fig. 5C–E) show that the bias favoring the sine fit persists up to a sampling rate of around 30Hz.

Discussion

Cognitive neuroscience has provided many insights into how spoken language is processed in the brain, yet we still lack a full understanding of how speech comprehension is implemented in the auditory system. Many studies have shown that speech is accompanied by oscillatory brain activity that phase aligns to the acoustic input, a mechanism that has been termed entrainment (e.g., Ref. [1,20]).

Transcranial electric stimulation has been suggested to entrain brain activity and it thus seems to be a well-suited tool to improve speech processing since it is engaging in the same mechanism. Recent work has indeed provided evidence that tES can improve spoken language comprehension using transcranial alternating current stimulation, tACS [10], as well as speech derived envelope-tES [6,7]. Therefore, it seems that tES targeting brain oscillations provides an effective tool to improve speech perception [21].

In the current study we tested whether pulsed-tES derived from transients in the speech material can interact with brain activity and improve speech perception similar to previous tACS and envelope-tES studies. We used tES to deliver brief electrical pulses to the auditory cortex at relevant times (syllable onsets) during spoken sentences and tested if this approach can improve perception as well. We show that subjects perform better in a speech comprehension task when pulsed-tES is applied with a critical delay (100 ms). This is in line with previous findings using envelope-tES [6] and falls within the time window of the N1, a strongly researched obligatory component of the evoked auditory cortical response (approximately 100 ms post-stimulus; e.g., Refs. [23], for a review). The auditory N1 has been connected to basic feature processing of acoustic input generated in the auditory cortex in the superior temporal cortex regions [24,25], which was the target of our tES montage. Thus, the behavioral benefit could be a result of improved acoustic feature processing of the speech material in the auditory cortex driven by the simultaneously arriving electric signal (since electricity travels instantaneously through the tissue, see below). Therefore, the electric pulses might have phase aligned ongoing brain activity such that the following syllable fell into an optimal cortical excitability state.

Furthermore, the auditory system is highly predictive, encoding temporal relationships of the acoustic input on different time scales (e.g., Ref. [26]) not only based on strictly rhythmic structures but also input from different modalities is integrated [27]. Along that line our results fit well with studies from cross-modal perception research. For instance, brief stimuli presented in the visual domain reset ongoing brain activity in auditory cortex, which in turn improves processing of the auditory input [28,29]. Critically, for maximal effects the visual stimuli must precede the auditory stimuli by a critical interval (~20–80 ms). Electric pulses might work similarly to incoming activation from a different modality, by adjusting the phase of ongoing activity. This way a better excitation phase for the auditory inflow is provided. However, since electrical activity modulates auditory cortex regions instantaneously, pulsed-tES does not need to precede auditory stimulation, which gives it an advantage for potential hearing-aid application.

We hypothesized that pulsed-tES, if applied with the correct timing, can improve speech perception. It is possible that, mechanistically, pulsed-tES improved entrainment of brain activity to the incoming acoustic speech signal. This might be especially so, since the speech signal we used was masked by noise, which degraded the acoustic features essential for entrainment (see Ref. [30]). Previous results using envelope-tES [6,7] could be explained by a similar mechanism. Even though envelope-tES contains the continuous quasi-rhythmic modulation of the signal envelope (cf. Fig. 1C with a clear peak in the spectrum), it also contains strong amplitude-increases typically in the beginning to center of a syllable. Therefore, it is possible that these strong signal peaks (due) to transients in the speech signal – comparable to the tES-pulses in the present work - drive the behavioral effects.

We tested our behavioral data for a sinusoidal modulation and found evidence for theta entrainment if looking at a reduced dataset (sampling one 4Hz theta phase every 50 ms, i.e., 20Hz sampling rate). However, this evidence is the result of a bias, and the analysis of the whole dataset (sampling one 4Hz cycle every 25 ms, i.e., 40Hz sampling) did not provide evidence for a phasic

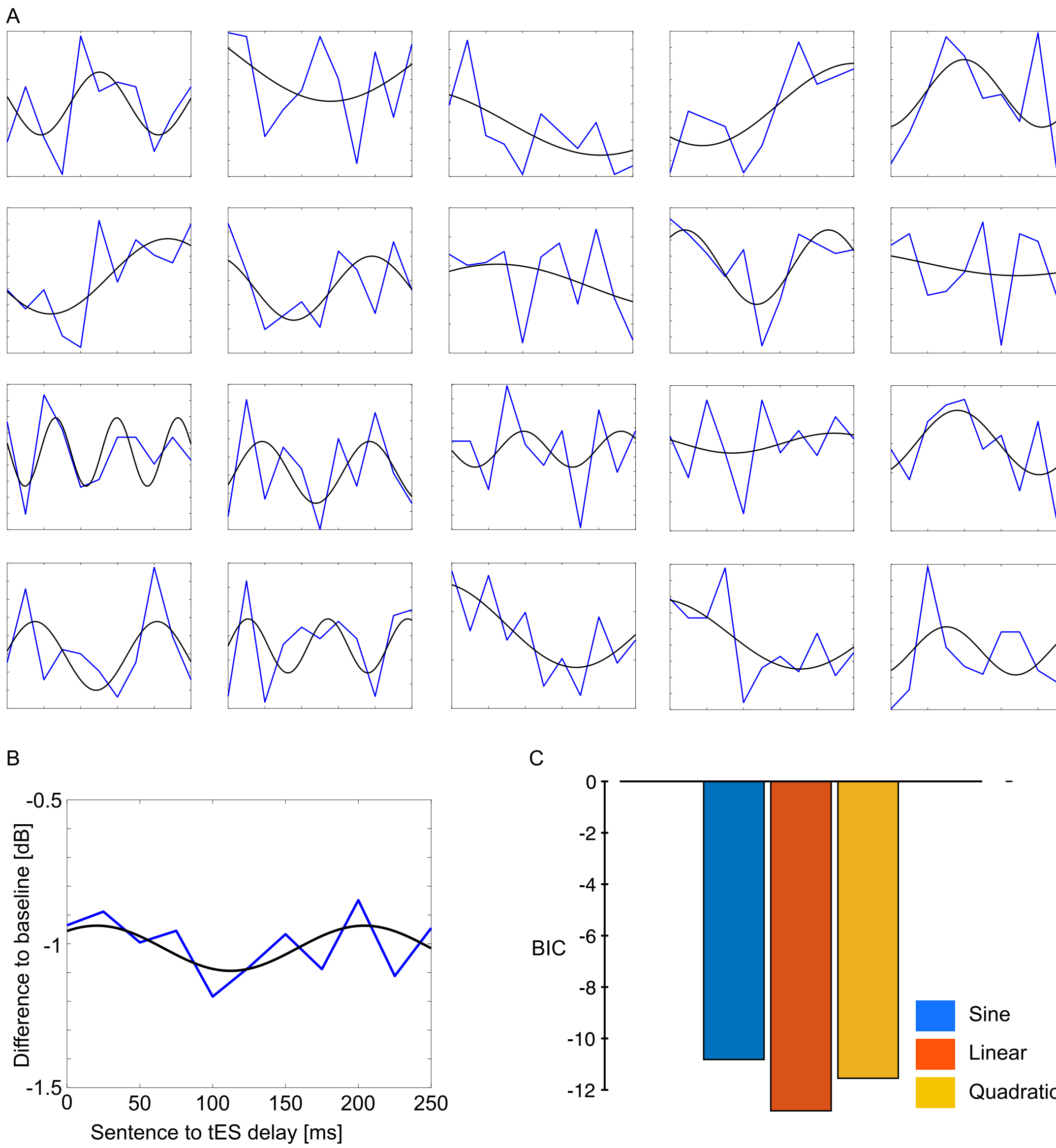


Fig. 3. (A) Sine fitting results for the individual subjects (B) and at the group level. (C) Bayes information criterion (BIC) for sinusoidal, linear, and quadratic fitting procedures. The BIC values are on a similar level. The individual as well as grand-average data descriptively show that a sine is not a superior fit (compared to the linear or quadratic model) in most cases (but see Fig. 4). Note that since BICs are compared descriptively no error bars are displayed.

modulation. This result clearly speaks against entrainment to the pulsed tES sequences by itself. It is possible that the non-oscillatory nature of the pulsed-tES sequences hinders entrainment of theta oscillations. Since it is already quite controversially discussed whether the quasi-rhythmicity of speech is sufficient to yield entrainment by itself (see Ref. [30]) this could be even more relevant for the - compared to the speech signal - highly distorted pulsed tES sequences. Even though the pulsed tES contains temporal regularities since it is based on the syllable structure (Fig. 1E), it does not contain a sinusoidal modulation and peak at the typical speech amplitude envelope frequency bands (cf. Fig. 1C–D). This heavy distortion might be the reason that pulsed tES can aid speech perception with the right timing (by potentially improving entrainment to the sentence), while not leading to an entrainment by itself.

We also used simulated data to characterize the effects of behavioral sampling rate on model fitting and reveal a clear bias for sinusoidal modulations (indicating entrainment) in 20Hz sampling rates which disappears in 40Hz sampling rates. When analyzing this further, we find that this bias remains strong until random data are sampled with at least 30 Hz (Fig. 5C–D). This emphasizes the importance of proper analysis tools and the investigation of biases in simulated data (see Fig. 5 and [9]) before conducting experiments. Especially in times where the scientific community is doubtful about results of electric brain stimulation (e.g., Ref. [31–33]) it is critical to validate the effectiveness of commonly used tES methods.

Our results should not be confined to research on transcranial electric stimulation. Another body of studies that investigates entrainment uses rhythmic sensory stimulation (e.g., light flashes and tones, for a review see Ref. [34]). There have been contradictory

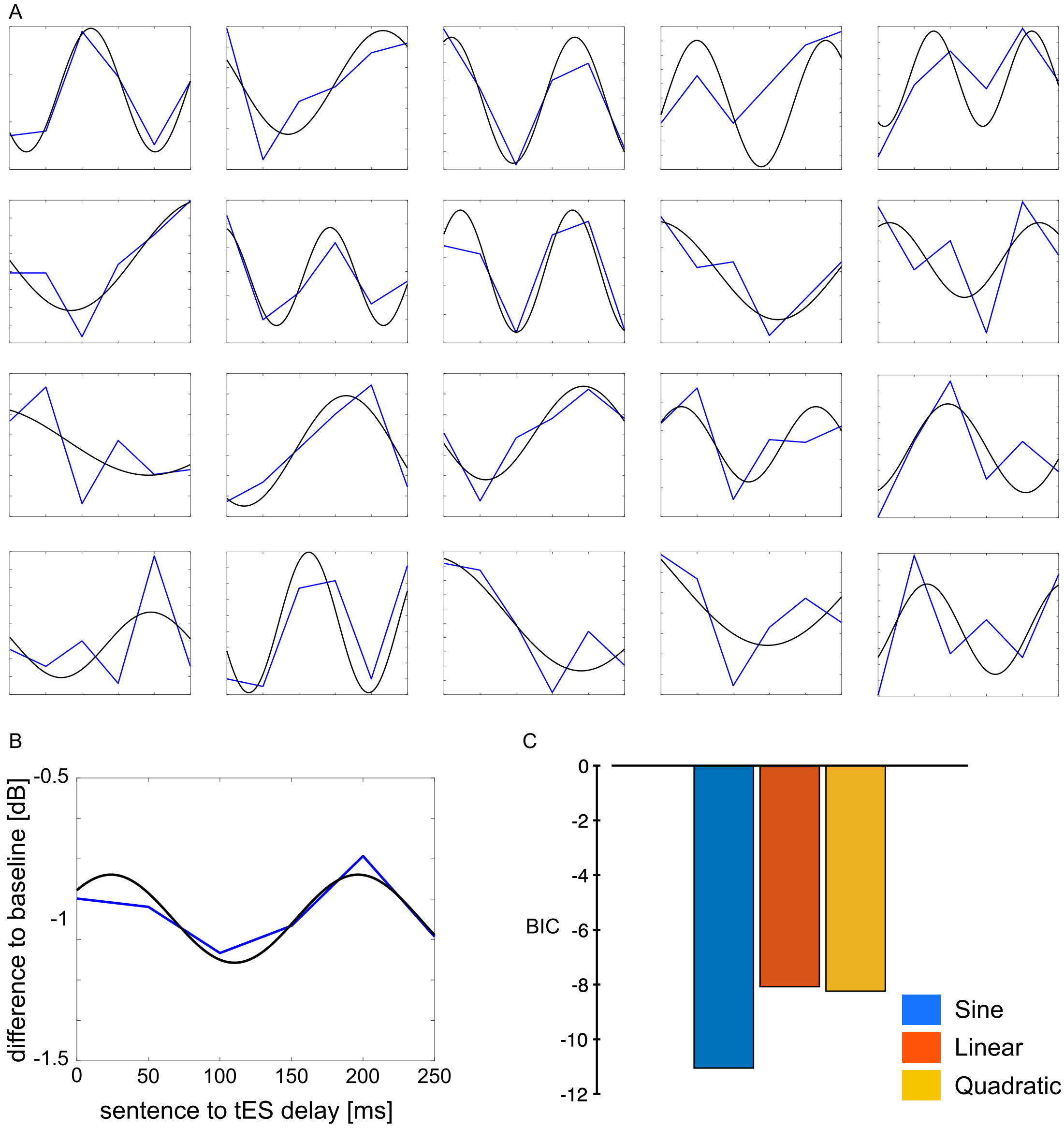


Fig. 4. Same as Fig. 3 but for half the data points (0–250 ms in 50 ms steps, 20Hz sampling rate) (A) Sine fitting results for the individual subjects and (B) for the group level. A sinusoidal modulation is clearly visible in the individual plots and grand-average. (C) Group average BIC results of the sinusoidal, linear, and quadratic fitting procedures. The sine fit wins over the other procedures. It thus seems likely that an oscillatory mechanism is captured by the low sampling rate data.

findings on the effectiveness of the entraining stimulus material as well (e.g., Refs. [35,36]). Our result can provide additional explanations for these conflicting results. A quadratic modulation of behavior driven by rhythmic task-irrelevant entrainers [35] was based on only 5 data points (non-equally sampled with a sampling rate between 18 and 47Hz). This approach can make a quadratic modulation a likely finding, which our simulations hint at. That is, a bias of quadratic vs. linear fit exists for low sampling frequencies below 25Hz (cf. Fig. 5E). This finding further underlines the importance of testing the analytic tools (e.g., via simulations) when investigating oscillatory entrainment.

Our study probably underestimates the individual benefit from the stimulation. Based on inter-individual differences in anatomy and cortical processing, the maximum efficiency of pulsed-tES is likely different across subjects (e.g., Refs. [37]), a circumstance often used as justification for selecting the maximum performance as a basis of the analysis. However, these inter-individual differences are challenging to account for without introducing a bias into the analyses [9].

A limitation to our study is a missing peripheral stimulation control condition. Recent work has shown that in certain cases, assumed cortical effects of non-invasive brain stimulation can be attributed to non-cortical sources, such as the retina [38] or somatosensory receptors [39]. To reduce sensory effects the stimulation intensity here was adjusted to sensation level (or slightly below). However, future studies should perform control conditions using, e.g., remote stimulation locations or analgesic cream blocking local receptors under the scalp [39] or compare the measured effects with individual electric fields simulation in areas of the brain and periphery [40]. A further limitation, is the single subject modeling that we apply. Adding experimental conditions (here the different temporal delays in that the tES-pulses were applied)

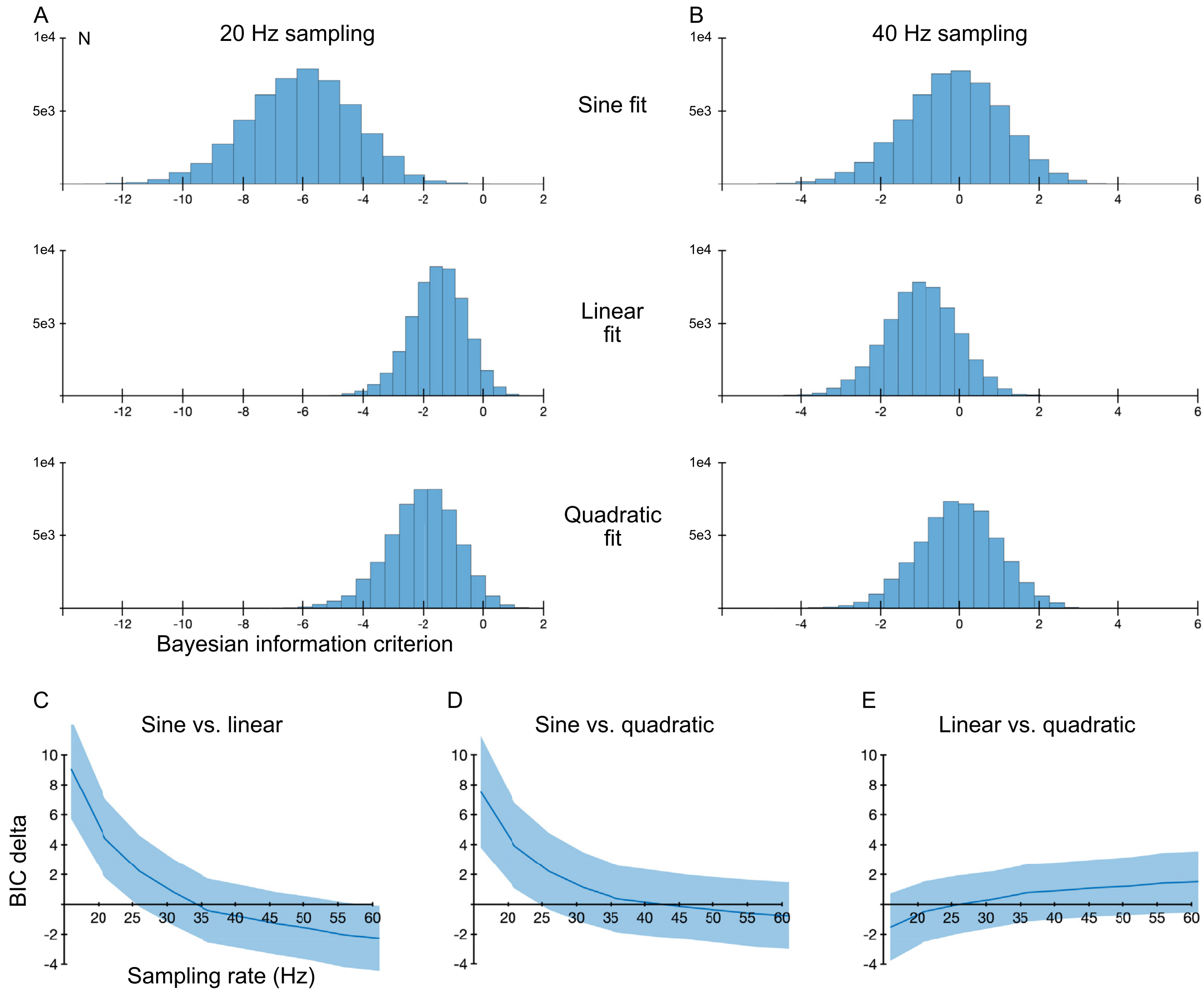


Fig. 5. Monte Carlo simulation results: Top - Comparison of BIC distributions estimated from fitting results of Monte Carlo simulated random data sampled at 20 Hz (A) and at 40 Hz (B). The (normal distributed) random data were simulated using the experimental sample data n, mean, and standard deviation. Monte Carlo simulation n = 10,000. The 20 Hz sampling creates a bias towards a sine fit, while there is no bias for the 40 Hz sampling rate. Bottom - Estimates of BIC bias (sine, linear, quadratic fit) using different sample sizes. BICs were estimated from Monte Carlo simulations (n = 5000) of random data with the parameters used as in A/B. Graphs show BIC differences of sine vs. linear (C), sine vs quadratic (D), and linear vs, quadratic (E) fits. Positive values indicate a bias towards sine fit (C/D) and linear fit (E). Negative values indicate a bias towards linear (C) and quadratic fit (D/E). Note that BIC differences smaller than 2 are not meaningful [20,21]. The shaded area represents 2 standard errors of the mean.

increases the noise level and could have thus obscured sinusoidal modulations of behavior. While the bias towards sine fits is clearly apparent with fewer conditions, the evidence against an improved model fit of the sinusoidal model compared to the linear and quadratic model has to be taken with a grain of salt.

In this study, we tested a novel protocol with the power to improve speech comprehension, which is computationally simple and energy efficient. Brief DC-pulses presented at around 100 ms after the detection of transients in the acoustic input (i.e., the start of a syllable) seems fairly easily accomplished in a real world setting with a hearing aid connected to a wearable electric stimulation device. Our findings indicate that there is no need for a pre-set frequency or signal processing of the acoustic input as used in other protocols.

In sum, we describe a pulsed tES-approach that can improve speech perception by delivering brief electrical pulses to the auditory system, potentially improving entrainment of auditory cortex activity to the acoustic input. Furthermore, we show that pulsed tES does not entrain brain activity itself, yet suboptimal data analysis would have led us to believe so. Future studies including electrophysiological measures are required to further elucidate the working mechanism behind pulsed-tES.

CRediT authorship contribution statement

Philipp Ruhnau: Data curation, Writing - original draft, planned the study. **Katharina S. Rufener:** Data curation, Writing - original draft, planned the study, collected the data. **Tino Zaehle:** Writing -

original draft, planned the study, All authors participated in interpreting the results, writing the manuscript and approved the final version.

Declaration of competing interest

The authors declare no competing financial interests.

Acknowledgments

The authors thank Tomko Settgast for help with the stimulus material preparation and data acquisition and Janika Wolter-Weging for help with data acquisition. This study was supported by the German Research Foundation (DFG ZA 626/2-1), the federal state of Saxony-Anhalt and the European Regional Development Fund (ERDF) in the Center for Behavioral Brain Sciences (CBBS, ZS/2016/04/78113). We thank 3 anonymous reviewers for their helpful comments on earlier versions of this manuscript.

References

[1] Giraud A-L, Poeppel D. Cortical oscillations and speech processing: emerging computational principles and operations. Nat Neurosci 2012 Mar;15(4):511–7.
[2] Ding N, Melloni L, Zhang H, Tian X, Poeppel D. Cortical tracking of hierarchical linguistic structures in connected speech. Nat Neurosci 2016 Jan;19(1):158–64.
[3] Ghitza O. The theta-syllable: a unit of speech information defined by cortical function. Front Psychol 2013;4:138.
[4] Gross J, Hoogenboom N, Thut G, Schyns P, Panzeri S, Belin P, et al. Speech rhythms and multiplexed oscillatory sensory coding in the human brain. PLoS Biol 2013 Dec;11(12):e1001752.
[5] Zoefel B, VanRullen R. EEG oscillations entrain their phase to high-level features of speech sound. Neuroimage 2015 Sep:1–8.
[6] Wilsch A, Neuling T, Obleser J, Herrmann CS. Transcranial alternating current stimulation with speech envelopes modulates speech comprehension. Neuroimage 2018 May;172:766–74.
[7] Riecke L, Formisano E, Sorger B, Başkent D, Gaudrain E. Neural entrainment to speech modulates speech intelligibility. Curr Biol 2018 Jan;28(2):161–169.e5.
[8] Riecke L, Sack AT, Schroeder CE. Endogenous delta/theta sound-brain phase entrainment accelerates the buildup of auditory streaming. Curr Biol 2015 Dec;25(24):3196–201.
[9] Asamoah B, Khatoun A, Laughlin MM. Analytical bias accounts for some of the reported effects of tACS on auditory perception. Brain Stimul Basic Transl Clin Res Neuromodulation 2019 Jul 1;12(4):1001–9.
[10] Zoefel B, Archer-Boyd A, Davis MH. Phase entrainment of brain oscillations causally modulates neural responses to intelligible speech. Curr Biol 2018 Feb;28(3):401–408.e5.
[11] Wagener K, Kühnel V, Kollmeier B. Entwicklung und Evaluation eines Satztests für die deutsche Sprache I: design des Oldenburger Satztests 1999;38(1):4–15.
[12] Wagener K, Brand T, Kollmeier B. Entwicklung und Evaluation eines Satztests für die deutsche Sprache Teil III: evaluation des Oldenburger Satztests. Z Audiol 1999;38(3):86–95.
[13] Heimrath K, Fiene M, Rufener KS, Zaehle T. Modulating human auditory processing by transcranial electrical stimulation. Front Cell Neurosci 2016;10:53.
[14] Kingdom FAA, Prins N. Psychophysics: a practical introduction [Internet]. Academic Press; 2016. Available from, https://books.google.com/books/about/Psychophysics.html?id=3sHQBAAAQBAJ.
[15] Neuling T, Rach S, Wagner S, Wolters CH, Herrmann CS. Good vibrations: oscillatory phase shapes perception. Neuroimage 2012 Nov;63(2):771–8.
[16] Schwarz G. Estimating the dimension of a model. Ann Stat 1978 Mar;6(2):461–4.
[17] Buckland ST. Monte Carlo confidence intervals. Biometrics 1984 Sep;40(3):811.
[18] Kass RE, Raftery AE. Bayes factors. J Am Stat Assoc 1995;90(430):773–95.
[19] Raftery AE. Bayesian model selection in social research 1995;25:111–63.
[20] Kösem A, Bosker HR, Takashima A, Meyer A, Jensen O, Hagoort P. Neural entrainment determines the words we hear. Curr Biol 2018 Sep;28(18):2867–75. e3.
[21] Zoefel B. Speech entrainment: rhythmic predictions carried by neural oscillations. Curr Biol 2018 Sep;28(18):R1102–4.
[23] Näätänen R, Picton T. The N1 wave of the human electric and magnetic response to sound: a review and an analysis of the component structure. Psychophysiology 1987 Jul;24(4):375–425.
[24] Key APF, Dove GO, Maguire MJ. Linking brainwaves to the brain: an ERP primer. Dev Neuropsychol 2005;27(2):183–215.
[25] Yvert B, Fischer C, Bertrand O, Pernier J. Localization of human supratemporal auditory areas from intracerebral auditory evoked potentials using distributed source models 2005;28(1):140–53.
[26] Schröger E, Marzecová A, SanMiguel I. Attention and prediction in human audition: a lesson from cognitive psychophysiology. Eur J Neurosci 2015 Mar;41(5):641–64.
[27] Talsma D. Predictive coding and multisensory integration: an attentional account of the multisensory mind. Front Integr Neurosci 2015;9(727):19.
[28] Kayser C, Petkov CI, Logothetis NK. Visual modulation of neurons in auditory cortex. Cerebr Cortex 2008 Jul;18(7):1560–74.
[29] Thorne JD, De Vos M, Viola FC, Debener S. Cross-modal phase reset predicts auditory task performance in humans. J Neurosci 2011;31(10):3853–61.
[30] Meyer L, Sun Y, Martin AE. Synchronous, but not entrained: exogenous and endogenous cortical rhythms of speech and language processing. Lang Cogn Neurosci 2019 Dec 11:1–11. 0(0).
[31] Liu A, Vöröslakos M, Kronberg G, Henin S, Krause MR, Huang Y, et al. Immediate neurophysiological effects of transcranial electrical stimulation. Nat Commun 2018 Nov;9(1):1–12.
[32] Vöröslakos M, Takeuchi Y, Brinyiczki K, Zombori T, Oliva A, Fernández-Ruiz A, et al. Direct effects of transcranial electric stimulation on brain circuits in rats and humans. Nat Commun 2018 Jan:1–17.
[33] Ruhnau P, Rufener KS, Heinze HJ, Zaehle T. Sailing in a sea of disbelief: in vivo measurements of transcranial electric stimulation in human subcortical structures. Brain Stimulat 2018;11(1):241–3.
[34] Haegens S, Golumbic EZ. Rhythmic facilitation of sensory processing_ A critical review. Neurosci Biobehav Rev 2018 Mar;86:150–65.
[35] Jones MR, Moynihan H, MacKenzie N, Puente J. Temporal aspects of stimulus-driven attending in dynamic arrays. Psychol Sci 2002 Jul;13(4):313–9.
[36] Bauer A-KR, Jaeger M, Thorne JD, Bendixen A, Debener S. The auditory dynamic attending theory revisited_ A closer look at the pitch comparison task. Brain Res 2015 Nov;1626(c):198–210.
[37] Rawji V, Ciocca M, Zacharia A, Soares D, Truong D, Bikson M, et al. tDCS changes in motor excitability are specific to orientation of current flow. Brain Stimulat 2018 Apr;11(2):289–98.
[38] Schutter DJLG, Hortensius R. Retinal origin of phosphenes to transcranial alternating current stimulation. Clin Neurophysiol Off J Int Fed Clin Neurophysiol 2010 Jul;121(7):1080–4.
[39] Asamoah B, Khatoun A, Laughlin MM. tACS motor system effects can be caused by transcutaneous stimulation of peripheral nerves. Nat Commun 2019 Jan 17;10(1):1–16.
[40] Kasten FH, Duecker K, Maack MC, Meiser A, Herrmann CS. Integrating electric field modeling and neuroimaging to explain inter-individual variability of tACS effects. Nat Commun 2019 Nov 28;10(1):1–11.